Inventory Control in Case of Unknown Demand and Control Parameters

Inventory Control in Case of Unknown Demand and Control Parameters

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 7 mei 2010 om 14.15 uur door

Elleke Janssen

geboren op 2 oktober 1980 te Nijmegen.

PROMOTOR: prof.dr.ir. D. den Hertog
COPROMOTOR: dr.ir. L.W.G. Strijbosch

Preface

> Het voorwoord is zowel in het Nederlands als in het Engels geschreven. Hieronder staat de Engelse versie; de Nederlandse versie kunt u vinden op pagina ix.

The preface is the place to thank all those people that have had a direct, or indirect, influence on the dissertation you are reading right now. I have chosen to write this part both in English and in Dutch, mainly because the mother tongue of the majority of the people mentioned in this preface is Dutch. However, since this dissertation is written in English, apart from the Dutch preface and summary in Dutch, also an English version of the preface should be present. This is not a direct translation of the Dutch one, but it covers the same topics.

Let me start thanking the person without whom this dissertation would never have appeared in the first place: Dick. When I was finishing my master's thesis, studying a topic that has nothing in common with this dissertation, he asked me to become a PhD-student. During my study Econometrics and OR I have never considered this career path, since doing purely theoretical research is not something I saw myself doing for four years. However, at the time I was finishing my master's thesis, there was a project at Tilburg University, in collaboration with Involvation, with a very practical application, namely inventory management. And they were looking for a PhD-student to join the project. So I started my career in academia. Dick is also my promotor, but not my daily supervisor; that role is for Leo, my copromotor. He helped me by brainstorming about the difficult theoretical parts in my dissertation, by giving me new ideas, but also by letting me follow up on my own ideas. I always liked going to our meetings, even when I was thinking "I did not get any step further since the last time, help!". In most cases our meeting let me see that, although I did not find the great idea on how to move forward, I at least had investigated, and rejected, some promising ideas that did not lead to anything. By letting me explain what I had done and what I had tried to show Leo helped me getting further. Even when Leo retired, approximately the last half year of writing this dissertation, he never shied away from reading my texts carefully and in doing so, he improved this dissertation significantly.

Also the discussions with and remarks of Ruud Brekelmans, Fred Janssen and Hans Moors have a great influence on the contents of this dissertation. They helped doing the research on which Chapter 2 (Hans), Chapters 3 and 4 (Ruud), and Chapter 5 (Fred) are based. One of these chapters (Chapter 3) is already published, a second one is almost published and Chapters 2 and 4 will be submitted in the near future.

The last group of people that has had a direct influence on the appearance of my dissertation is the committee that has read and, probably even more important, approved it. The committee consists of John Boylan, Ton de Kok, Fred Janssen, Ruud Brekelmans, Leo Strijbosch and Dick den Hertog.

Next to the relative small group of people that has had a direct influence, there is a rather large group that (very) indirectly influenced my dissertation. My colleagues at the department of Econometrics and Operations Research at Tilburg University belong to this group. They made sure that I liked going to my work all four years and that I still do. At almost all my working days I am looking forward to the lunch, not just because I was having an appetite, but mostly because of the talks on all kinds of serious and mostly non-serious topics, like the food in the university restaurant, the sport performances (or failures) of last weekend, television programmes, the political party one has voted for (or is going to vote for), the mall in Tilburg, the departmental trips that were going to happen or happened, the television in the Triangel and thousand-and-one other topics. In random order this involves Peter, Herbert, Henk, Hans Reijnierse, Jacob, Willem, Bart, Annemiek, Ruud Hendrickx, Gerwald, John, Gijs, Mark Voorneveld, Marieke, Marloes, Salima, Edwin Lohmann, Edwin van Dam, René, Hein, Ruud Brekelmans, Josine and, recently, Mirjam and Iris. After eating our lunch there is often some time left to kick back and relax a little more and that time we enjoy in the Triangel. Besides continuing the talks at lunch, watching sports (unfortunately, the television is not working anymore) and collecting items one gets when doing grocery shopping at Appie, we solve puzzles (cryptic crosswords, Mona), play games (Duck rally, Fluxx) or finding (three)double animals: words consisting of two (or three) animals, that have a non-animal meaning.

Aside from working on my dissertation and relaxing also some ‘real’ work needed to be done. I taught several courses (Statistics for HBO-graduates, Mathematics 2, Quantitative Methods 1, Statistics 1, BEM, Mathematics 1 for Economics, Mathematics D) and I would like to thank Marieke, Herbert, Willem, Edwin van Dam, Gert and Marloes for the nice cooperation during these courses. During the first two years of working at my dissertation I found out that I like teaching very much and when I got the opportunity to extend my contract in exchange for extra education tasks, I took that opportunity. I should thank Marieke, Herbert, Peter and Sprint, the people (and organization) that made this extension possible.

My officemates at Tilburg University, Marlies, Marloes, Katya, Gijs, Frans, Roy and Mohammed, always had (and have) time for a chat, a cup of coffee, jokes on radio Veronica or adventure stories of last weekend in which the police does not always have a positive role. Luckily they also knew when to let my work. Many thanks to all of you.

Also outside the university buildings there is time to relax with (part of) my colleagues during Christmas dinners, mathematics D dinners, departmental trips, visits to the Efteling, Sinterklaas and mostly the game afternoons and evenings, the (jigsaw) puzzle evenings and the subdepartmental trips (Ruud Hendrickx, Marieke, John, Gerwald, Edwin Lohmann, Marloes, Josine, Gijs, Mirjam and Iris).

A special paragraph should be awarded to Marieke, Salima and Marloes. Next to all the fun things we do, these three girls have been there for me during difficult (personal) times. And they have given me a push when I needed it.

One of the advantages of working as a PhD-student are the conference visits. I have met nice people (Ingrid Vliegen and her colleagues of TU Eindhoven, Marco Bijvank, Rommert Dekker, Ruud Teunter, John Boylan and Aris Syntetos), listened to interesting presentations (and even more not so interesting presentations…) and enjoyed visiting new and interesting cities.

Besides all colleagues and (other) people from academia, I also want to thank everyone outside this world that has supported me. My parents, Wim and Annelies, have always believed that my talents made me special, they have made me do my best, but have never pushed me in one direction and let me find out what I liked doing. Who would have imagined that I would end up in education, just like my father…

Raike and Matthijs, my sister and brother-in-law, and Rob, my brother, might not have had a big influence, but just knowing that they are there for me, helps me. When I went to the university, $10\frac{1}{2}$ years ago, and I came home for the weekend, the first thing my brother used to ask me, was "When are you going back?", but nowadays we can have good talks. The long telephone calls with my sister help me keeping a hold on reality, since that could get a little lost in academia.

To my family — my grandmother, my aunts and uncles, my cousins — who still think of me being a regular student: I have finished and starting a 'real' job, although I still will be working at the university for a couple more years. Their questions about what I was doing exactly, were often difficult to answer without being (to) technical. My four-year research resulted in this book (most likely unreadable for those without a mathematics background) and my doctor's title (if everything goes like planned at May, 7th).

Also Ab and Ada, my parents-in-law and Frank, Jozien and Lars, my brother-in-law, sister-in-law and nephew, have always supported me by being there for Mark and me. We have had many nice conversations, cosy diners, some nice holidays and celebrated Sinterklaas and Christmas together. They helped my, together with my family, to keep in touch with reality and made sure that I did not start thinking too much of myself just because I am able to learn a little easier than the average person. The birth of nephew opened up a new world for me and showed me how miraculously fast a human being can develop and grow.

Last, but certainly not least, I must thank Mark, who is already almost nine years my boyfriend, my friend and my love. He is there for me when I need him. Sometimes just by putting his arm around me, by offering me a shoulder to cry on or by lending me his ear; often by enjoying nature together, by planning to travel and making these plans come true, by having nice and beautiful holidays or just by eating together.

Voorwoord

> The preface is available both in English and in Dutch. This is the Dutch version; you can find the English version at page v.

Het voorwoord is de plek waar menigeen bedankt wordt voor de direct of indirecte invloed op het tot stand komen van het proefschrift dat nu voor u ligt. Ik heb ervoor gekozen om dit stukje van mijn proefschrift zowel in het Nederlands als in het Engels te schrijven, omdat Nederlands de moedertaal is voor verreweg de meeste mensen die een plekje hebben weten te veroveren in dit voorwoord. Aangezien de rest van het proefschrift, op de samenvatting na, in het Engels is, is het voorwoord ook in het Engels beschikbaar en alhoewel het geen letterlijke vertaling is, staat er wel hetzelfde in.

Laat ik beginnen met degene te bedanken zonder wie dit proefschrift überhaupt nooit geschreven zou zijn: Dick. Toen ik bezig was met het onderzoek dat zou leiden tot mijn afstudeerscriptie, over een onderwerp dat overigens totaal niets met mijn proefschrift te maken heeft, heeft Dick mij gevraagd om AiO te worden. Ik heb tijdens mijn studie altijd gezegd dat dat niets voor mij was, dat continu bezig zijn met (theoretisch) onderzoek. Nu was er op dat moment een project met praktische toepassing, namelijk voorraadbeheer, in samenwerking met Involvation. En bij dat project was plek voor een AiO, waarbij Dick aan mij dacht. Zo ben ik dus mijn AiO-schap ingerold. Dick is ook mijn promotor, maar de ‘dagelijkse’ begeleiding lag meer bij mijn copromotor, Leo. Hij heeft me geholpen door mee te denken over moeilijke stukken in mijn onderzoek, nieuwe ideeën aan te dragen, maar mij ook mijn eigen dingen te laten doen. Onze afspraken waren nooit iets om tegen op te zien, alhoewel ik wel eens dacht “Ik ben geen *** opgeschoten, wat nu?”. Maar dan bleek tijdens onze afspraken dat ik stiekem toch wel iets verder gekomen was, doordat ik uit moest leggen wat me allemaal niet gelukt was. Zelfs toen Leo al met pensioen was, ongeveer het laatste halve jaar, heeft hij de tijd genomen om alle teksten secuur te blijven lezen en daardoor nog een flink aantal verbeteringen aangebracht.

Naast Leo heb ik inhoudelijk ook veel gehad aan de opmerkingen van en discussies met Ruud Brekelmans, Fred Janssen en Hans Moors. Ze hebben samen met Leo en mij het onderzoek gedaan waarop hoofdstuk 2 (Hans), hoofdstukken 3 en 4 (Ruud) en hoofdstuk 5 (Fred) gebaseerd zijn. Ze zullen dus ook co-auteur zijn van de papers die nog gaan komen, of die al (bijna) gepubliceerd zijn.

Wat betreft de directe bijdrage aan mijn proefschrift behoort nog één groep mensen bedankt te worden: de commissie die dit proefschrift gelezen heeft en, misschien wel belangrijker, het goedgekeurd heeft. Dit zijn John Boylan, Ton de Kok, Fred Janssen, Ruud Brekelmans, Leo Strijbosch en Dick den Hertog.

Naast de relatief kleine groep die direct invloed heeft gehad op dit proefschrift is er een grote groep die een (zeer) indirecte invloed heeft uitgeoefend. Hiertoe behoren mijn collega's die ervoor gezorgd hebben dat ik het al die jaren bij het departement Econometrie & Operations Research naar mijn zin heb gehad (en nog steeds heb). Zo goed als iedere dag dat ik gewerkt heb, heb ik uitgekeken naar de lunch. En niet zozeer omdat ik trek had of zo, maar vanwege de gezellige en vaak ook onzinnige gesprekken aan tafel over het eten in de mensa, de verscheidene sportprestaties (of wanprestaties) van het afgelopen weekend, boer zoekt vrouw, daten in het donker en andere televisieprogramma's, de politieke partij waarop je gaat stemmen (of gestemd had), de Tilburgse mall, de departementsuitjes die eraan kwamen of geweest waren, de televisie in de Triangel en nog duizend-en-één andere onderwerpen. In willekeurige volgorde gaat het hierbij om Peter, Herbert, Henk, Hans Reijnierse, Jacob, Willem, Bart, Annemiek, Ruud Hendrickx, Gerwald, John, Gijs, Mark Voorneveld, Marieke, Marloes, Salima, Edwin Lohmann, Edwin van Dam, René, Hein, Ruud Brekelmans, Josine en sinds kort ook Mirjam en Iris. Na de lunch is het tijd voor nog iets meer ontspanning om over de after-lunch-dip heen te komen en daarvoor is de Triangel uitermate geschikt gebleken. Naast een vervolg op de gesprekken aan de lunchtafel, het volgen van een of andere wedstrijd of het sparen van producten die bij de Appie te krijgen zijn, is hier altijd ruimte voor een puzzeltje (cryptogrammen, Mona), een spelletje (eendenrally, Fluxx) of het vinden van (drie)dubbeldieren, zoals zebrapad, vlinderdas en kadodoos.

Uiteraard moet er ook nog eens af en toe gewerkt worden en naast het onderzoek voor en schrijven aan mijn proefschrift heb ik behoorlijk wat onderwijs gegeven. Daarin heb ik zeer prettig samengewerkt met Marieke, Herbert, Willem, Edwin van Dam, Gert en Marloes. Ik kwam er in mijn eerste twee jaar achter dat ik onderwijs eigenlijk erg leuk vind en toen ik de kans kreeg om mijn contract met een jaartje te verlengen in ruil voor extra onderwijs, heb ik die kans met beide handen aangegrepen, met dank aan Marieke, Herbert, Peter en Sprint, die dat mogelijk gemaakt hebben.

Bij mijn kamergenoten op de UvT, Marlies, Marloes, Katya, Gijs, Frans, Roy en Mohammed, was (en is) er altijd tijd voor een praatje, een kop koffie, moppen op radio Veronica of een verhaal over het afgelopen weekend, waar dan de politie een niet altijd positieve hoofdrol heeft. Gelukkig wisten ze ook wanneer ze me aan het werk moesten laten gaan. Veel dank daarvoor.

Ook naast het werk op de UvT is er de nodige tijd om te ontspannen met (een deel van) mijn collega's tijdens kerstdiners, wiskunde D etentjes, departementsuitjes, bezoek aan de Efteling, Sinterklaasvieringen en met name de spelletjesavonden, de puzzelavonden en de subdepartementale uitjes (Ruud Hendrickx, Marieke, John, Gerwald, Edwin Lohmann, Marloes, Gijs, Josine, Mirjam, Iris).

Een speciale alinea moet besteed worden aan Marieke, Salima en Marloes, die er altijd voor me zijn geweest om me door moeilijke (privé-)momenten heen te helpen, naar me te luisteren, een arm om me heen te slaan en me dat zetje te geven dat ik nodig had. En dat naast alle leuke dingen die we samen doen en gedaan hebben.

Eén van de voordelen van het werken als AiO zijn de (betaalde) congresbezoeken. Ik heb daarbij leuke mensen ontmoet (met name, maar niet enkel, Ingrid Vliegen en haar collega's van de TU Eindhoven, Marco Bijvank, Rommert Dekker, Ruud Teunter, John Boylan en Aris Syntetos), interessante praatjes aangehoord (en nog meer minder interessante praatjes. . .) en interessante steden bezocht.

Naast alle collega's en (andere) academici, zijn er nog meer mensen die me vooral moreel gesteund hebben. Mijn ouders, Wim en Annelies, die altijd geloofd hebben dat mijn talenten mij bijzonder maken en me altijd gesteund hebben om zover mogelijk te komen, maar me nooit in een richting gestuurd hebben en me zelf hebben laten uitzoeken wat ik wilde. Dat ik, net als mijn vader, uiteindelijk in het onderwijs terecht zou komen, zou ik tot een paar jaar geleden niet bedacht hebben. . .

Raike en Matthijs, mijn zusje en mijn zwager, en Rob, mijn broertje, hebben misschien inhoudelijk niet veel bijgedragen, maar het feit dat ze er zijn als het nodig is, is voor mij genoeg. Toen ik ging studeren, tien-en-een-half jaar geleden, en ik in het weekend thuis kwam, was het eerste dat mijn broertje tegen me zei: "Wanneer ga je weer?", maar tegenwoordig kunnen we goede gesprekken hebben. De lange telefoongesprekken met mijn zusje zijn altijd goed om me weer met beide benen in de werkelijkheid te zetten.

Aan de rest van mijn familie — mijn oma, ooms, tantes, neven, nichten, en achterneefjes en -nichtjes — voor wie ik toch een beetje de eeuwige student ben: ik ben nu klaar en ga 'echt' werken, al blijf ik nog wel een tijdje op de universiteit hangen. Hun vragen over wat ik nu precies aan het doen was, waren vaak moeilijk te beantwoorden zonder al te veel in technische details te duiken en toch te laten blijken dat het niet zo eenvoudig was als dat ik meestal vertelde. Mijn onderzoek, dat vier jaar geduurd heeft, heeft uiteindelijk geleid tot dit, voor niet-wiskundigen waarschijnlijk onleesbare, boekje en tot de titel van 'doctor' (als alles goed gaat op 7 mei 2010).

Ook mijn schoonouders, Ab en Ada, mijn zwager en schoonzus, Frank en Jozien, en mijn neefje Lars, hebben me altijd gesteund door er voor mij en Mark te zijn. We hebben vele fijne gesprekken gehad, vaak gezellig gegeten, soms samen op vakantie geweest en Sinterklaas en Kerst samen gevierd. Samen met mijn eigen familie, zorgen zij ervoor dat ik het nooit te hoog in mijn bol heb gekregen doordat ik toevallig wat beter dan gemiddeld kan leren. Mijn neefje heeft een hele nieuwe wereld voor mij geopend en laat zien hoe wonderbaarlijk snel een mens(je) kan groeien, zowel figuurlijk als letterlijk.

Last, but certainly not least, moet ik Mark bedanken, die al bijna negen jaar mijn vriend, maatje en liefste is. Hij is er altijd als ik hem nodig heb. Soms door een arm om me heen te slaan, door een schouder te bieden om op uit te huilen of een luisterend oor te geven; vaak door samen te genieten van de natuur, samen (plannen te maken om) te reizen, mooie vakanties te hebben of gewoon samen te eten.

Contents

Chapter 1

Introduction

This chapter starts with a background literature review. The next section contains notation that is used throughout this dissertation. Section 1.3 introduces inventory control as it is used in this research. The last section provides an overview of this dissertation.

1.1 Background

Inventory control involves decisions on what to order, when, and in what quantity. Standard text books on inventory management (see e.g., Silver et al., 1998, or Zipkin, 2000) provide methods to deal with these decisions. These methods need information about the (distribution of) demand during some period, e.g., the demand during lead time or during the review period. Bulinskaya (1990) discriminates between three situations:

(a) the type of distribution is known, but its parameters are unspecified;

(b) only several first moments of the demand distribution are known;

(c) there is no prior knowledge about the demand.

The third situation is of course the most realistic and different approaches to deal with situation (c) have been proposed in literature. These approaches can be categorized into parametric and nonparametric methods. An example of a parametric method is using Bayesian models; examples of the nonparametric methods include using order statistics, the bootstrap procedure and kernel densities.

One of the most widespread approaches to deal with unknown demand is assuming a distribution, estimating its parameters and replacing the unknown parameters by its estimates in the theoretically correct formulae in which distribution and parameters are supposed to be known. Sani and Kingsman (1997), Artto and Pylkkänen (1999), Strijbosch et al. (2000) and Syntetos and Boylan (2006) use this approach with different inventory models, while Kottas and Lau (1980) provide a short discussion on estimating the parameters needed for their model. Another parametric method is the Bayesian approach; Azoury and Miller (1984), Azoury (1985) and Karmarkar (1994) are three examples of this approach. Also Larson et al. (2001) use it, but they introduce a nonparametric form. Other nonparametric approaches involve order statistics, references include Lordahl and Bookbinder (1994) and Liyanage and Shanthikumar (2005), the bootstrap procedure, see, e.g., Bookbinder and Lordahl (1989) and Fricker and Goodhart (2000), or using kernel densities, see Strijbosch and Heuts (1992).

This dissertation is mainly about the effect of forecasting on inventory control. Most of the literature is either on forecasting or on inventory control, but one can easily imagine that forecasting influences inventory control. Although not many papers consider both forecasting and inventory control, the problem has already been mentioned in 1958 (Scarf, 1958). He considers situation (b): it is assumed that the mean and variance of demand are known and considers a set of two-point distributions to solve a max-min objective function (maximize the minimal profit). Hayes (1969) considers situation (a) with two different demand distributions. More recent references include Watson (1987), Strijbosch and Heuts (1992), Snyder et al. (2002), Bertsimas and Thiele (2006), Lu et al. (2006) and Syntetos and Boylan (2006). Watson (1987) considers Erlang distributed demand and studies the effect forecasting has on the attained service using simulation. Strijbosch and Heuts (1992) use simulation to show the trade-off between attained service and expected average costs while estimating the lead time demand in four different ways, including a distribution-free approach. Snyder et al. (2002) use simulation to show the effects of using adapted exponentially smoothed forecasts, which incorporates the possibility of having nonconstant variance. Bertsimas and Thiele (2006) assume that the mean and variance of the demand are known, while the family to which it belongs, is not and use robust optimization to find good inventory control parameters. Lu et al. (2006) focus on the way the demand forecasts evolve over time as more information becomes available and use that to find solution bounds and cost error bounds for general dynamic inventory models with possibly nonstationary and autocorrelated demands. Syntetos and Boylan (2006) compare four estimators for intermittent demand and study their stock control performance using an empirical data sample containing monthly demand of 3000 stock keeping units during a period of two years.

Part II of this dissertation deals with the effect of estimating unspecified parameters. Therefore, it deals with situation (a), not (c): the true distribution is known, but its parameters are unspecified. Silver and Rahnama (1986, 1987) investigate the effect of estimating parameters in a reorder point, order quantity inventory policy (known as an (s, Q) or (r, Q) policy in literature) with a cost criterion. They construct a function that determines the expected cost of estimating the demand distribution rather than knowing it and they conclude that this function is not symmetrical: underestimating causes larger costs than overestimating. In the second article they propose a method that deliberately biases the reorder point upwards. Strijbosch et al. (1997) and Strijbosch and Moors (2005) investigate the same effect for a periodic review, order-up-to level inventory policy with a service level criterion under normally distributed demand. Both papers conclude that also in this case the order-up-to level needs to be biased upwards. Note that the reorder point, order quantity and periodic review, order-up-to level policies are equivalent (see Silver et al., 1998), so the results of Silver and Rahnama (1986, 1987) apply for the periodic review, order-up-to level policy and the results of Strijbosch et al. (1997) and Strijbosch and Moors (2005) for the reorder point, order quantity policy as well.

Both in practice and in literature demand is often assumed to be normally distributed, see Zeng and Hayya (1999). However, the normal distribution can take on negative values, while in practice demand cannot be negative (unless one considers net demand: demanded goods minus returned goods). Strijbosch and Moors (2006) develop two modified normal distributions to tackle this problem: one is truncated at zero; the other assigns a value of zero to all negative values, creating a point mass at zero. They derive the new order-up-to levels and show comparisons between results using the new demand distributions and mistakenly using a (nonmodified) normal distribution. One could see this as a particular case of situation (b), as the wrong distribution is assumed, while the mean and variance of the demand are known.

In Part I of the dissertation these modified demand distributions are discussed further and two new demand distributions are introduced. A gamma distribution does not have negative values, but one can shift such a distribution to the left, and then negative values will occur. Starting from this shifted gamma distribution we develop two modified shifted gamma distributions.

1.2 Notation

This section contains the notation as it is used throughout this dissertation. If in the chapter itself some specific notation is used, it is explained in that chapter.

D_ℓ	: demand during ℓ units of time
D	: demand during 1 unit of time
R	: length of the review period
L	: length of the lead time
S	: order-up-to level
α	: P_1 service level; cycle service
β	: P_2 service level; fill rate
$\mathbb{P}\left(A\right)$	: probability of event A
$\mathbb{E}\left[X\right], \mu_X$	: expected value of X
$\mathbb{V}\left[X\right], \sigma_X^2$	: variance of X
$\mathbb{SD}\left[X\right], \sigma_X$	: standard deviation of X ($\mathbb{SD}\left[X\right] = \sqrt{\mathbb{V}\left[X\right]}$)
$\mathbb{C}_X, \nu_X$	: coefficient of variation of X ($\mathbb{C}_X = \nu_X = {}^{\sigma_X}/_{\mu_X}$)
$(x)^+$	: maximum of 0 and x ($(x)^+ = \max\{x, 0\}$)
$\mathcal{I}_C$	: indicator function: 1 if C is true and 0 otherwise
pdf	: shorthand notation for probability density function
cdf	: shorthand notation for cumulative distribution function
$f(x)$	: a (general) probability density function (pdf)
$F(x)$	: a (general) cumulative distribution function (cdf)
$f_{\text{parameters}}(x)$	: a pdf with its distribution parameters
$F_{\text{parameters}}(x)$	: a cdf with its distribution parameters
$\varphi(x)$	: the pdf of a standard normally distributed variate
$\Phi(x)$	: the cdf of a standard normally distributed variate

The length of the review period, length of the lead time and the order-up-to level are denoted with capital letters; this usually implies that these variables are random variables. However, it depends on the chapter whether these are random or not. The order-up-to level is assumed to be random in Part II; the length of the review period and lead time are only assumed to be random in Chapter 5.

1.3 Introduction to the (R, S) policy and service levels

This section provides a short introduction to the (R, S) inventory control policy and service levels as it is used throughout this dissertation. We have chosen the (R, S) policy, because derivations and calculations are relatively easy. Hence, we can use this policy to illustrate our results, which often involve analytical derivations. Furthermore, the results we obtain for the (R, S) policy also hold for the (s, Q) policy, since the (R, S) and (s, Q) policy are equivalent (Silver et al., 1998). Extending the results to other inventory control policies is one direction for further research.

1.3.1 Inventory control policy

Inventory is needed for selling products (inventory of final products) or for producing products (inventory of materials and semi-finished products). However, inventory should not be too high, as holding inventory costs money. So there is a trade-off between service offered and inventory costs. With help of an inventory control policy such a trade-off can be made in a thoughtful manner.

This dissertation considers a periodic review, order-up-to level inventory control policy. The (R, S) policy reviews the inventory position every R periods and replenishes it up to the order-up-to level S. That order is delivered L periods later. The inventory position at time t (IP_t) is not only the inventory on hand at that time (OH_t), but also considers the orders that are not yet delivered (the pipeline inventory, PI_t) and the demand that is not yet satisfied (the backlog, BL_t). If a customer arrives and the inventory is not sufficient to satisfy its demand, that demand is backlogged; i.e., the demand is satisfied as soon as new products are delivered by the firm's supplier. The net stock at time t (NS_t) is defined as $OH_t - BL_t$, so if positive, inventory is present ($OH_t = NS_t$) and there is no backlog ($BL_t = 0$). If NS_t is negative, then no inventory is available at the firm ($OH_t = 0$) and some demand is backlogged ($BL_t = -NS_t$). The inventory position is determined according to

$$IP_t = OH_t - BL_t + PI_t = NS_t + PI_t. \tag{1.1}$$

The order size at time t, Q_t, is determined by $Q_t = S - IP_t$. Figures 1.1 and 1.2 display these different terms graphically for two cases: $R > L$ and $R < L$. In Figure 1.1 the first order (Q_0) is placed at time 0. This order raises the inventory position

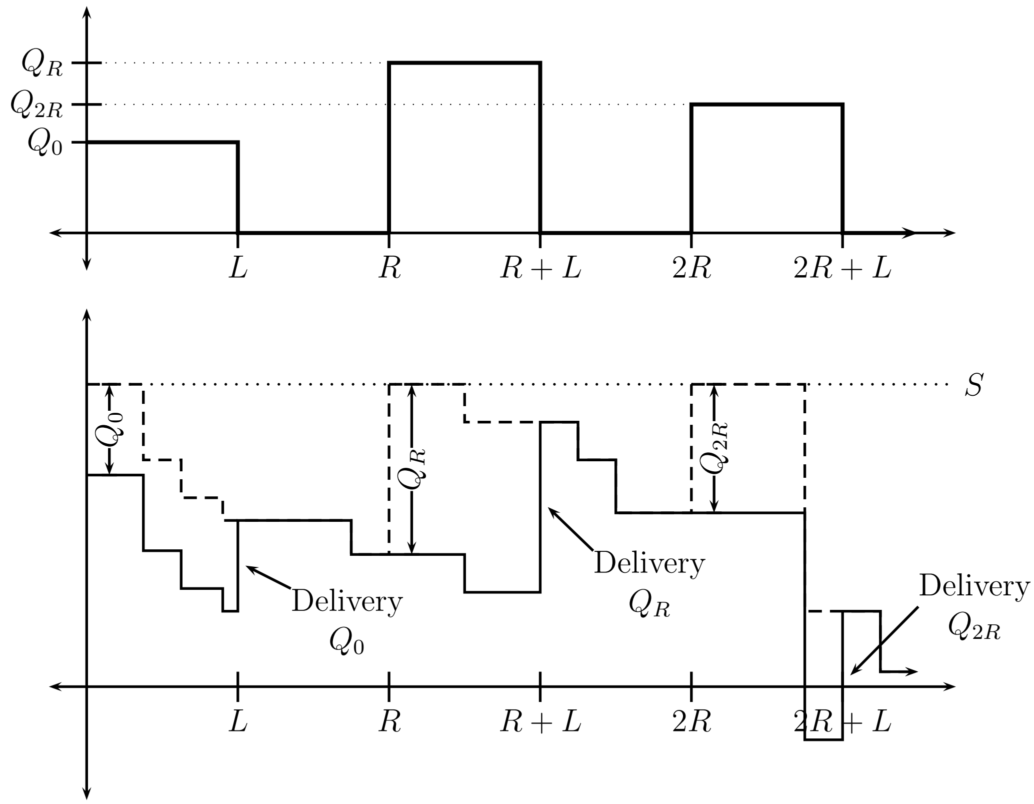


Figure 1.1: Inventory position (dashed line), net stock (solid line) and pipeline inventory (thick solid line) displayed graphically for $R > L$.

up to S, but it does not change the net stock. The amount of orders in the pipeline is now Q_0. Then some customers arrive and their demands are satisfied. These demands lower both the inventory position and the net stock. At time L the order placed at time 0 arrives; this raises the net stock and lowers the amount of orders in the pipeline both by Q_0. The difference between net stock and inventory position is the amount of orders in the pipeline and since no orders are left in the pipeline, the inventory position and net stock coincide. Then, at time R, the cycle of order placement and order delivering starts again. Note that shortly prior to the delivery of the third order (Q_{2R}) the net stock drops below 0, which means that the firm is out of stock and part of the demand of the customer arrived at that time, is backlogged. It is satisfied when the order Q_{2R} is delivered. Finally, note that the order sizes differ for different time epochs. This policy has a fixed order-up-to level and fixed time between orders, hence the amount ordered can vary. For the (s, Q) policy the contrary holds: the order sizes are equal at each order epoch (it is Q) but the time

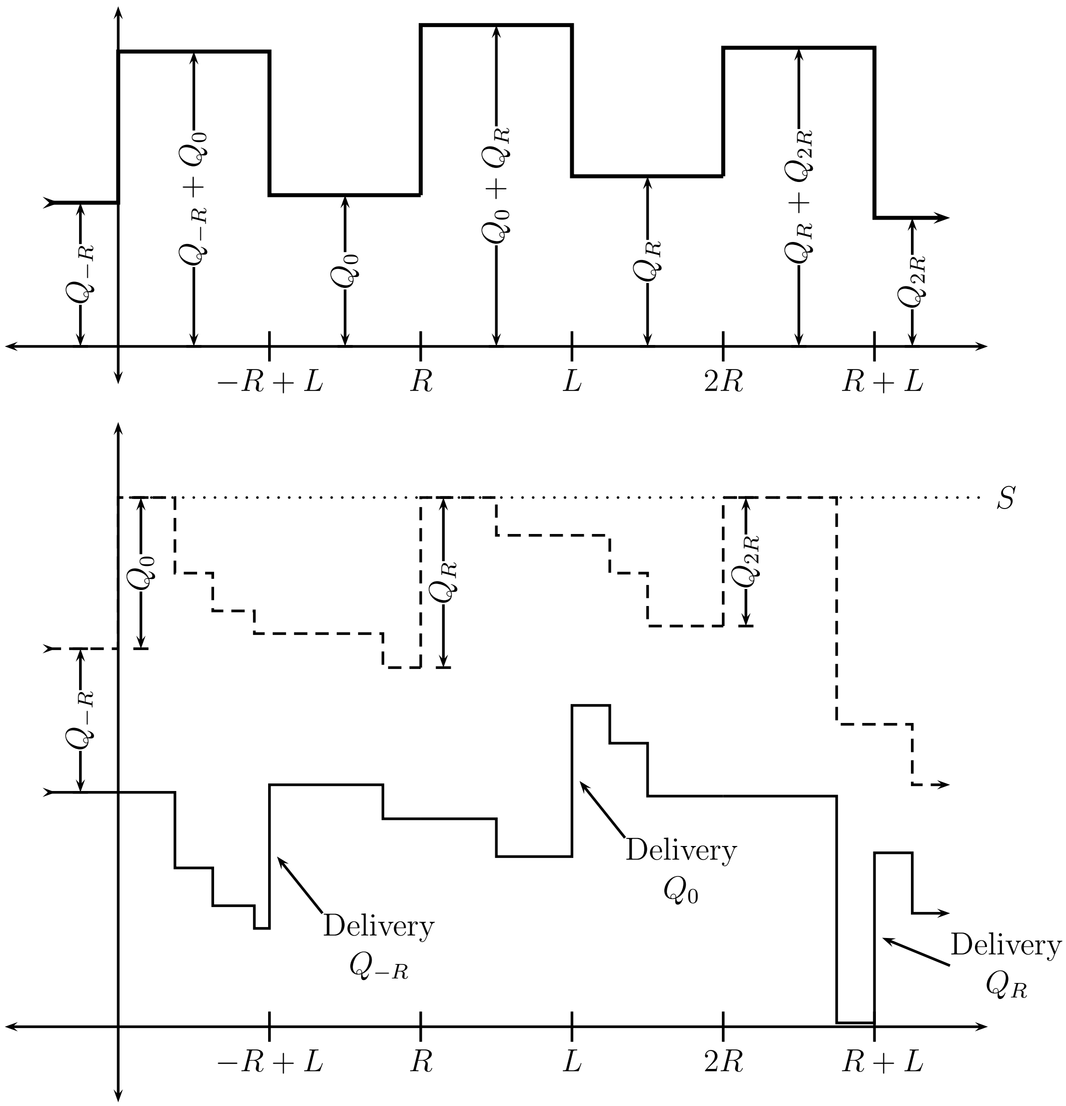


Figure 1.2: Inventory position (dashed line), net stock (solid line) and pipeline inventory (thick solid line) displayed graphically for $R < L$ (order Q_{-R} is placed at the last review before time 0).

between orders and the inventory position just after ordering vary.

In Figure 1.2 the lead time is longer than the review period. Just before time 0 one order is in the pipeline. That order is placed at the last review before time 0, so at time $-R$. At time 0 a new order is placed (Q_0). This raises the inventory position and the amount of orders in the pipeline by Q_0, so now there are two orders in the pipeline. After the review moment customers arrive and their demands are satisfied from stock. At time $-R+L$ the first order, placed at time $-R$, arrives. This raises the net stock by Q_{-R} and lowers the pipeline inventory by the same amount. Now only the order placed at time 0 is left in the pipeline. Again a customer arrives and his demand can be satisfied from stock. At time R a new order is placed, another customer arrives and then, at time L, the order placed at time 0 is delivered. Note that in this graph the inventory position and net stock never coincide, since there is always at least one order in the pipeline. If the probability of no demand during a review period is negligible, the net stock and inventory position never coincide under an (R, S) policy with $L > R$.

The value of the order-up-to level S should be chosen in such a way that the firm can make a good trade-off between service offered and inventory costs incurred. The preferred method for making this decision is minimizing costs, but one needs costs for backlogging demand or for lost sales and those costs are usually extremely difficult to determine (see, e.g., Silver et al., 1998). Hence, service levels are used instead. We impose that S needs to be large enough to attain a certain service and by assuming that this service level is attained exactly (and not exceeded) inventory is kept as low as possible. Hence, inventory costs are not too high and the desired service is reached. How to choose the value of the order-up-to level exactly is subject of this dissertation.

1.3.2 Service level

In order to decide whether the service the firm provides to its customers is good enough, one has to define it. This dissertation considers two types of service levels:

- P_1 service level;
- P_2 service level.

Both types consider stock out occurrences during the replenishment cycle. A replenishment cycle is the time between two order deliveries. Note that a replenishment cycle has length R (see Figures 1.1 and 1.2). At the start of the replenishment cycle the net stock is equal to the order-up-to level minus the demand during the lead time: $S - D_L$. At the end, just before the next delivery, it equals $S - D_{R+L}$. Note that S is fixed, while D_L and D_{R+L} are random variables.

P_1 service level

The P_1 service level measures the fraction of replenishment cycles without stock out occurrences. In literature it is also referred to as cycle service. Both terms are used in this dissertation, as well as the notation α. Mathematically, the P_1 service level is defined as:

$$\alpha = \mathbb{P}\left(D_{R+L} \leq S\right). \tag{1.2}$$

Example 1.1 (P_1 service)
Assume that demand during review plus lead time is distributed according to:

d_{R+L}	15	25	35	45
$\mathbb{P}\left(D_{R+L} = d_{R+L}\right)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

If we set the order-up-to level to 25 ($S = 25$), the P_1 service is

$$\alpha_{S=25} = \mathbb{P}\left(D_{R+L} \leq 25\right) = \frac{4}{8} = 0.50.$$

If we set the order-up-to level to 35, the P_1 service is

$$\alpha_{S=35} = \mathbb{P}\left(D_{R+L} \leq 35\right) = \frac{7}{8} = 0.8750.$$

□

P_2 service level

The P_2 service level measures the fraction of demand satisfied directly from shelf. In literature it is often referred to as fill rate. Both terms are used in this dissertation, as well as the notation β. Mathematically, the P_2 service level is defined as:

$$\beta = 1 - \frac{\mathbb{E}\left[(D_{R+L} - S)^+\right] - \mathbb{E}\left[(D_L - S)^+\right]}{\mathbb{E}\left[D_R\right]}. \tag{1.3}$$

The second term in the numerator prevents counting backlog twice: if there is already a backlog at the start of the replenishment cycle, this should be subtracted from the backlog at the end of the replenishment cycle. This second term is not always taken into account in literature, mostly since the probability of having backlog at the start of a replenishment cycle is negligible if the desired service level is high, which is common in practice.

Example 1.2 (P_2 service)
Assume that demand during lead time is distributed according to

d_L	5	15
$\mathbb{P}(D_L = d_L)$	$\frac{1}{2}$	$\frac{1}{2}$

the demand during the review period according to

d_R	10	20	30
$\mathbb{P}(D_R = d_R)$	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

and the demand during the review plus lead time according to

d_{R+L}	15	25	35	45
$\mathbb{P}(D_{R+L} = d_{R+L})$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

If we set the order-up-to level to 25 ($S = 25$), the P_2 service is determined by

$$\begin{aligned}\mathbb{E}\left[(D_{R+L} - 25)^+\right] &= (35-25)\cdot\frac{3}{8} + (45-25)\cdot\frac{1}{8} = \frac{50}{8}\\ \mathbb{E}\left[(D_L - 25)^+\right] &= 0\\ \mathbb{E}[D_R] &= 10\cdot\frac{1}{4} + 20\cdot\frac{2}{4} + 30\cdot\frac{1}{4} = 20\\ \beta_{S=25} &= 1 - \frac{\frac{50}{8} - 0}{20} = \frac{11}{16} = 0.6875.\end{aligned}$$

If we set the order-up-to level to 35, the P_2 service is

$$\begin{aligned}\mathbb{E}\left[(D_{R+L} - 35)^+\right] &= (45-35)\cdot\frac{1}{8} = \frac{10}{8}\\ \mathbb{E}\left[(D_L - 35)^+\right] &= 0\\ \mathbb{E}[D_R] &= 10\cdot\frac{1}{4} + 20\cdot\frac{2}{4} + 30\cdot\frac{1}{4} = 20\\ \beta_{S=35} &= 1 - \frac{\frac{10}{8} - 0}{20} = \frac{15}{16} = 0.9375.\end{aligned}$$

$\square$

One can think of many more service level definitions, for example the fraction of time the net stock is positive or the probability that an arbitrary customer has to wait. This dissertation focuses on the P_1 and P_2 service levels. One cannot say that one of the two is better without knowing anything about the product and its market. Consider the following two examples.

Example 1.3 (P_1 versus P_2 service (1))
A shop sells office supplies. The demand for a set of pens in the past ten days is given below.

day	1	2	3	4	5	6	7	8	9	10
demand	90	92	88	91	88	86	93	89	91	92

The store has an (R, S) policy, with the review period equal to one day ($R = 1$) and the order-up-to level equal to 90 ($S = 90$). Its lead time is 0: the order is made after closing and delivered the next morning before opening again.

The attained P_1 service level in the past 10 days is only 0.50, while the attained P_2 service level is 0.99. In this case one could argue that the P_2 service level is more suitable, since the shopkeeper is more interested in the amount of demand he can satisfy directly. He does not mind too much whether he can satisfy all the demand during one day or not, as long as he satisfies most of the demand. □

Example 1.4 (P_1 versus P_2 service (2))
A maintenance department of a big factory has inventory of spare parts. The demand for a certain spare part in the past ten days is given below.

day	1	2	3	4	5	6	7	8	9	10
demand	7	11	8	12	9	9	13	9	12	10

The store has an (R, S) policy, with the review period equal to one day ($R = 1$) and the order-up-to level equal to 9 ($S = 9$). Its lead time is 0: the order is made after the end of the last shift and delivered the next morning before the first shift starts.

The attained P_1 service level in the past 10 days is only 0.50, while the attained P_2 service level is 0.91. In this case one could argue that the P_1 service level is more suitable, since it is important that all the demand in a replenishment cycle is satisfied. If a certain spare part is not available, a machine does not work and a factory cannot produce, hence the factory does not generate revenue. □

So the P_1 service level is best suited for situations in which missing one item is just as bad as missing multiple items, e.g., spare parts. The P_2 service suits situations in which the amount of nonsatisfied demand is important, e.g., retail stores and wholesale business.

1.4 Contributions and overview

The remainder of this dissertation is split into two parts. The first part, consisting of Chapter 2, deals with modified demand distributions. This is an example of situation (b), since we assume that we do know the mean, variance and third moment of demand. In this part the demand follows a non-standard distribution and we show the effect of using a wrong distribution (fitted with help of the mean, variance and third moment of the demand) on the achieved performance. The second part, consisting of Chapters 3–5, considers situation (a): not knowing the exact characteristics of demand. In classic inventory control demand distribution is assumed to be known completely. In practice this is rarely true. A demand distribution is assumed and its parameters are estimated using historical demand observations. The effect of replacing true, but unknown, parameters by their estimates is subject of the second part. Chapter 6 concludes this dissertation.

Chapter 2 defines two modified shifted gamma distributions to be used in inventory control. It provides a method to find the order-up-to levels under the new demand distribution and compares the results to the regular and shifted gamma distribution.The main contribution of this chapter lies in providing demand distributions that are more flexible.

Chapter 3 considers demand with a normal distribution with unknown parameters. Under strong assumptions it is shown analytically that the desired service level is not met when using estimates instead of the true parameters. When these assumptions are relieved, analytical derivations are no longer possible, but simulation shows that also now the desired service is not met. This chapter provides a method that improves the attained service and assures that the desired service is (almost) met. The method is based on the analytical proof of not reaching the desired service and on a regression technique. The main contributions of this chapter are the analytical proof of not reaching the desired service level, and the development of a correction function.

Chapter 4 considers demand with a gamma distribution with unknown parameters. Also in this case the desired service level is not met when using estimates. This is shown analytically under strong conditions and with simulation if these conditions are relieved. A method that improves the attained service, based on analytical derivations and regression, is provided. Simulation shows that the desired service level is (almost) met. The main contributions in this chapter are the analytical proof of not reaching the desired service level, and the development of the correction functions.

Chapter 5 is more general compared to Chapters 3 and 4. An effect of estimating the demand parameters is that the order-up-to level becomes a random variable (see Sections 3.2 and 4.2.1). This chapter takes this given as a starting-point: the order-up-to level is a random variable instead of a fixed number, as are the length of the review period and the length of the lead time. The order-up-to level, the demand during the review period and the demand during the lead time follow a mixed-Erlang distribution. It is shown that the desired service level is not met when ignoring the randomness. The correct order-up-to levels are derived analytically. The main contribution in this chapter is simultaneously considering demand, the length of the review period, the length of the lead time, and the order-up-to level to be random.

Chapter 6 provides the overall conclusion and directions for further research.

Part I

Modified distributions

Chapter 2

Modified shifted gamma distributions

This chapter considers the shifted gamma distribution: the gamma distribution that also has a location parameter. If we assume that demand is distributed according to this gamma distribution, negative demand observations may occur, which is often not realistic in practice. The two modified shifted gamma distributions discussed in this chapter only have nonnegative values. Assuming that demand is distributed according to the modified shifted gamma distribution, the order-up-to level in an (R, S) inventory control policy is derived under the P_1 and under the P_2 service level constraint. Finally, the results are compared to using a regular or shifted gamma distribution, while the true distribution is one of the two modified shifted gamma distributions.

2.1 Introduction

The normal distribution is often used to model demand, because of its tractability (see also Sections 1.1 and 3.1). The two main problems with this distribution are that it is symmetric, and the probability of negative realizations is non-negligible (more than 1%) if the coefficient of variation is larger than 0.43. In practice, demand is often skewed to the right and, more important, negative demand hardly ever occurs. Strijbosch and Moors (2006) have solved these two problems by constructing two modified normal distributions: one with a point mass at zero and one truncated at zero. These two modified normal distributions are skewed to the right and cannot have negative realizations. They discuss these new distributions and their limitations (the coefficient of variation cannot be large). Furthermore, they use these distributions to model the demand in an (R, S) inventory control policy with cycle service and fill rate constraints. Finally, they show that using the regular normal distribution while in fact the demand is distributed according to a modified normal distribution leads to underperformance. The modified normal distribution with a point mass at zero has the nice side-effect that it can be used as the distribution of intermittent demand, since the probability of zero demand is positive. Intermittent demand is often modeled using a compound distribution: the demand occurrences follow some distribution, e.g., the Bernoulli distribution, and the demand sizes follow another distribution, e.g., the gamma distribution; see, e.g., Janssen et al. (1998). Thus, we need two distributions to model the intermittent demand if we use the compound distribution approach, whereas one of the modified normal distributions captures the intermittency directly.

Another way to capture the nonnegativity of demand is using the gamma distribution to model demand (see also Section 4.1). This distribution does not have negative realizations and, furthermore, it is skewed to the right. Also the gamma distribution is easy to work with, although it does not have all the nice properties of the normal distribution. A disadvantage is that the probability of having zero demand is zero, and therefore this distribution cannot be used for intermittent demand.

Starting from the regular gamma distribution one can construct a shifted gamma distribution, simply by shifting the complete pdf either to the left or to the right. If it is shifted to the left, negative realizations can occur and we can use this shifted gamma distribution to construct two modified shifted gamma distributions, analogously to Strijbosch and Moors (2006).

These two modified shifted gamma distributions have the major advantage that they are more flexible than the regular gamma distribution, since the modified shifted gamma distribution has three parameters instead of the two parameters of the regular gamma distribution. Furthermore, the modified shifted gamma distribution with a point mass at zero can be used to directly model intermittent demand, since the probability of having zero demand is positive.

These two modified shifted gamma distribution are used to model demand during review in an (R, S) inventory control policy with lead time equal to zero. So, every R periods the inventory position is replenished up to the order-up-to level S and that order is delivered instantaneously. The size of the order-up-to level is chosen such that either the cycle service or the fill rate reaches a prescribed value. Note that the assumption of zero lead time is not as rigid as it might seem at first sight. Consider, e.g., a supermarket that makes its order at the end of opening hours and receives this order the next morning before opening hours. The time it takes the supplier to deliver is then approximately 12 hours (depending on the opening hours), but since the supermarket is closed during that time, one can take the lead time equal to zero.

This chapter starts with a summary of the results obtained by Strijbosch and Moors (2006) (the use of modified normal distributions to model demand). Next, the shifted gamma distribution is defined and also the two modified shifted gamma distributions are constructed from this shifted gamma distribution. In Section 2.4 the order-up-to levels under the modified shifted gamma distributions are derived and some general results considering the order-up-to levels using modified distributions are provided. Section 2.5 considers the use of the regular and modified gamma distribution, while demand actually follows a modified gamma distribution. This chapter is concluded in Section 2.6.

2.2 Modified normal distribution

Strijbosch and Moors (2006) have introduced two modified normal distributions, that only consider nonnegative observations. They start from the regular normal distribution with pdf $f_{\mu,\sigma^2}(x)$ and cdf $F_{\mu,\sigma^2}(x)$.

The first modified normal distribution is obtained by setting the value of negative realizations of the regular normal distribution to 0. This leads to a point mass at zero with value $F_{\mu,\sigma^2}(0) = \Phi(-1/\nu)$, where $\Phi(x)$ is the cdf of the standard normal distribution. The pdf and cdf of this modified normal distribution, denoted by $f^+_{\mu,\sigma^2}(x)$ and $F^+_{\mu,\sigma^2}(x)$, are

$$f^+_{\mu,\sigma^2}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{\mu,\sigma^2}(0) & \text{if } x = 0 \\ f_{\mu,\sigma^2}(x) & \text{if } x > 0, \end{cases} \quad \text{and} \quad F^+_{\mu,\sigma^2}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{\mu,\sigma^2}(x) & \text{if } x \geq 0. \end{cases}$$

Figure 2.1 depicts $f^+_{\mu,\sigma^2}(x)$. The mean, variance and coefficient of variation as a function of σ and $\nu = \frac{\sigma}{\mu}$ are

$$\begin{aligned} \mu^+ &= \sigma G(-\tfrac{1}{\nu}), \\ \sigma^{+^2} &= \sigma^2\big(H(\tfrac{1}{\nu}) - G^2(-\tfrac{1}{\nu})\big), \\ \nu^+ &= \sqrt{\frac{H(\frac{1}{\nu})}{G^2(-\frac{1}{\nu})} - 1}, \end{aligned}$$

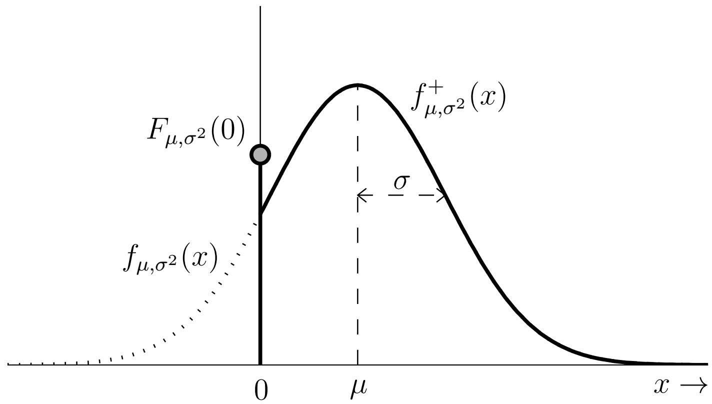


Figure 2.1: The pdf of a modified normal distribution with a point mass at zero (f^+_{μ,σ^2}).

where $G(x)$ denotes the loss function of a standard normal variate, i.e.,

$$G(x) = \mathbb{E}\left[(Z-x)^+\right] = \int_x^\infty (z-x)\varphi(z)dz = \varphi(x) - x\Phi(-x),$$

and $H(x)$ is an auxiliary function, which denotes

$$H(x) = x\varphi(x) + (x^2+1)\Phi(x).$$

Finally, $\varphi(x)$ and $\Phi(x)$ are the pdf and cdf of a standard normal variate.

The second modified normal distribution is constructed by ignoring all negative realizations of a regular normal distribution; this is a normal distribution truncated at zero. The pdf and cdf ($f^*_{\mu,\sigma^2}(x)$ and $F^*_{\mu,\sigma^2}(x)$) of this truncated normal distribution are

$$f^*_{\mu,\sigma^2}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{f_{\mu,\sigma^2}(x)}{\Phi\left(\frac{1}{\nu}\right)} & \text{if } x \geq 0, \end{cases} \quad \text{and} \quad F^*_{\mu,\sigma^2}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{F_{\mu,\sigma^2}(x) - \Phi\left(-\frac{1}{\nu}\right)}{\Phi\left(\frac{1}{\nu}\right)} & \text{if } x \geq 0. \end{cases}$$

The pdf belonging to the truncated normal distribution is depicted in Figure 2.2. The mean, variance and coefficient of variation belonging to this modified distribution, are

$$\begin{aligned} \mu^* &= \sigma \frac{G(-\frac{1}{\nu})}{\Phi(\frac{1}{\nu})}, \\ {\sigma^*}^2 &= \sigma^2 \frac{\Phi(\frac{1}{\nu})H(\frac{1}{\nu}) - G^2(-\frac{1}{\nu})}{\Phi^2(\frac{1}{\nu})}, \\ \nu^* &= \sqrt{\frac{\Phi(\frac{1}{\nu})H(\frac{1}{\nu})}{G(-\frac{1}{\nu})} - 1}. \end{aligned}$$

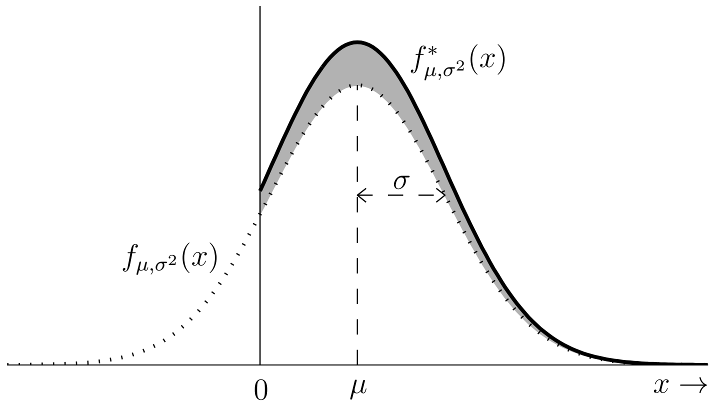


Figure 2.2: The pdf of a modified normal distribution truncated at zero (f^*_{μ,σ^2}).

Strijbosch and Moors (2006) provide expressions for the mean, variance and coefficient of variation of the modified normal distributions depending on σ and ν; the expressions for the coefficients of variation solely depend on ν. Furthermore, they show that the coefficients of variation (ν^+ and ν^* for the modified normal distribution with point mass at zero and the truncated normal distribution, respectively) are limited. They show that

$$\nu^+ \leq \lim_{\nu\to\infty} \nu^+ = \sqrt{\pi - 1} \approx 1.4634$$

and

$$\nu^* \leq \lim_{\nu\to\infty} \nu^* = \sqrt{\tfrac{\pi}{2} - 1} \approx 0.7555.$$

They use these modified distributions to model demand in an (R, S) inventory control policy with zero lead time under the cycle and fill rate service level constraints. This leads to expressions for the order-up-to levels: S^+_α (in case of P_1 service) and S^+_β (in case of P_2 service) for the modified normal distribution with a point mass at zero, and S^*_α (P_1 service) and S^*_β (P_2 service) for the truncated normal distribution. Using the loss function ($G(x)$) and the inverse of the cdf of the modified normal distribution ($\Phi^{-1}(x)$), Strijbosch and Moors (2006) derive

$$\begin{aligned}
S_\alpha &= \mu + \sigma\Phi^{-1}(\alpha), \\
S_\beta &= \mu + \sigma G^{-1}\left(\tfrac{1-\beta}{\nu}\right), \\
S^+_\alpha &= F^{+^{-1}}_{\mu,\sigma^2}(\alpha), \\
S^+_\beta &= \mu + \sigma G^{-1}\left(\tfrac{(1-\beta)\mu^+}{\sigma}\right),
\end{aligned}$$

$$S^*_\alpha = \mu + \sigma\Phi^{-1}\left(\alpha + (1-\alpha)\Phi\left(-\tfrac{1}{\nu}\right)\right),$$
$$S^*_\beta = \mu + \sigma G^{-1}\left(\frac{(1-\beta)\mu^*\Phi\left(\frac{1}{\nu}\right)}{\sigma}\right),$$

where S_α and S_β are the well-known order-up-to levels using a regular normal distribution. Using the order-up-to levels as defined above and the expressions for μ^+ and μ^*, it is easily shown that

$$S^+_\alpha = S_\alpha \qquad \text{and} \qquad S^*_\beta = S^+_\beta.$$

Furthermore, it holds that

$$S^*_\alpha > S_\alpha \qquad \text{and} \qquad S^+_\beta < S_\beta.$$

Section 2.4 shows that these properties also hold for the modified gamma distributions and, in fact, for general modified distributions constructed in an analogous way (Section 2.4.3).

Finally, Strijbosch and Moors (2006) show the consequences of using a regular normal distribution while in fact the demand is distributed according to the modified normal distribution: the desired service level is not reached and the larger the coefficient of variation is, the larger the underperformance is.

2.3 Modified shifted gamma distributions

This chapter considers two modified shifted gamma distributions, that start from the shifted gamma distribution. This distribution is also known as a Pearson type III distribution or a three-parameter gamma distribution, although there are also references to other three-parameter gamma distributions, so the latter is not uniquely defined. The shifted gamma distribution has three parameters: a location parameter Δ, a shape parameter ρ and a scale parameter θ. Its pdf ($f_{\rho,\theta,\Delta}(x)$) and cdf ($F_{\rho,\theta,\Delta}(x)$) are defined as

$$f_{\rho,\theta,\Delta}(x) = \begin{cases} \dfrac{(x+\Delta)^{\rho-1}e^{-\frac{x+\Delta}{\theta}}}{\theta^\rho\Gamma(\rho)} & \text{if } x \geq -\Delta \\ 0 & \text{if } x < -\Delta \end{cases} = \begin{cases} f_{\rho,\theta}(x+\Delta) & \text{if } x \geq -\Delta \\ 0 & \text{if } x < -\Delta, \end{cases}$$
$$F_{\rho,\theta,\Delta}(x) = \begin{cases} F_{\rho,\theta}(x+\Delta) & \text{if } x \geq -\Delta \\ 0 & \text{if } x < -\Delta, \end{cases}$$

where $f_{\rho,\theta}(x)$ and $F_{\rho,\theta}(x)$ denote the pdf and cdf of a regular gamma distribution (the shifted gamma distribution with $\Delta = 0$) and $\Gamma(x)$ denotes the gamma function. Figure 2.3 displays the pdf of a shifted gamma distribution and a regular gamma distribution with the same shape and scale parameter. Note that, in order to obtain

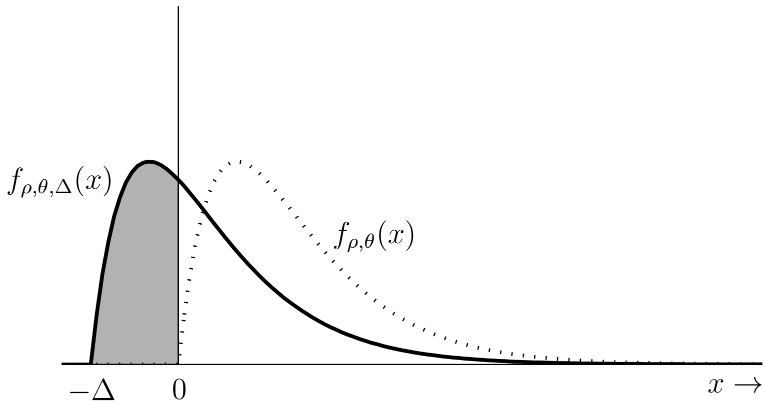


Figure 2.3: The pdf of a shifted gamma distribution, with probability of negative realizations depicted in gray, compared to the regular gamma distribution

the pdf of the shifted gamma distribution, the complete pdf of the regular gamma distribution with the same scale and shape parameters shifts Δ to the left. The mean and variance of a regular gamma distribution are $\rho\theta$ and $\rho\theta^2$; the mean and variance of the shifted gamma distribution are easily obtained using the shift of size Δ:

$$\begin{aligned}\mu &= \rho\theta - \Delta,\\ \sigma^2 &= \rho\theta^2,\\ \nu &= \frac{\sigma}{\mu} = \frac{\sqrt{\rho}\theta}{\rho\theta - \Delta}.\end{aligned}$$

Two important results from the regular gamma distribution can be translated to the shifted gamma distribution:

$$\begin{aligned}G_{\rho,\theta,\Delta}(x) &= \int_x^\infty (z-x) f_{\rho,\theta,\Delta}(z)dz\\ &= \rho\theta[1 - F_{\rho+1,\theta,\Delta}(x)] - (x+\Delta)[1 - F_{\rho,\theta,\Delta}(x)], \qquad (2.1)\\ x f_{\rho,\theta,\Delta}(x) &= \theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta f_{\rho,\theta,\Delta}(x). \qquad (2.2)\end{aligned}$$

The derivations of (2.1) and (2.2) are in Appendix A.1. Using (2.2), we obtain the following two expressions:

$$x^2 f_{\rho,\theta,\Delta}(x) = \theta^2\rho(\rho+1) f_{\rho+2,\theta,\Delta}(x) - 2\Delta\theta\rho f_{\rho+1,\theta,\Delta}(x) + \Delta^2 f_{\rho,\theta,\Delta}(x), \tag{2.3}$$

$$\begin{aligned} x^3 f_{\rho,\theta,\Delta}(x) = {} & \theta^3\rho(\rho+1)(\rho+2) f_{\rho+3,\theta,\Delta}(x) - 3\Delta\theta^2\rho(\rho+1) f_{\rho+2,\theta,\Delta}(x) \\ & + 3\Delta^2\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta^3 f_{\rho,\theta,\Delta}(x). \end{aligned} \tag{2.4}$$

Equation (2.3) is obtained by multiplying both the left-hand side and right-hand side of (2.2) by x and using (2.2) on the right-hand side again. Equation (2.4) is derived by multiplying both the left-hand side and right-hand side of (2.3) by x and using (2.2) on the right-hand side again. The exact derivations are also in A.1.

In general, the location parameter Δ could have any value, but since we start from a distribution with possibly negative realizations, Δ is assumed to be positive in the remainder of this chapter.

2.3.1 Modified gamma distribution with point mass at zero

The modified gamma distribution with a point mass at zero is obtained be setting the value of each negative realization of the shifted gamma distribution to zero. The graph of the pdf of such a distribution is displayed in Figure 2.4. The pdf and cdf

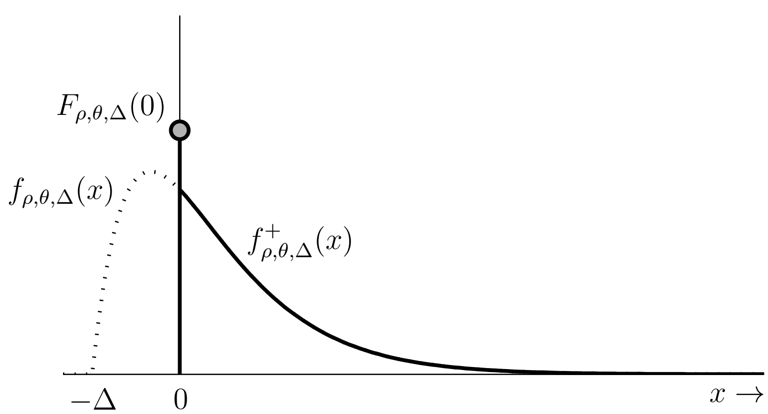


Figure 2.4: The pdf of the modified gamma distribution with a point mass at zero.

are defined as

$$f^+_{\rho,\theta,\Delta}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{\rho,\theta,\Delta}(0) & \text{if } x = 0 \\ f_{\rho,\theta,\Delta}(x) & \text{if } x > 0, \end{cases} \tag{2.5}$$

$$F^+_{\rho,\theta,\Delta}(x) = \begin{cases} 0 & \text{if } x < 0 \\ F_{\rho,\theta,\Delta}(x) & \text{if } x \geq 0. \end{cases} \tag{2.6}$$

The mean of this distribution, denoted by μ^+, is

$$\mu^+ = \int_0^\infty x f^+_{\rho,\theta,\Delta}(x)dx = 0 \cdot F_{\rho,\theta,\Delta}(0) + \int_0^\infty x f_{\rho,\theta,\Delta}(x)dx = G_{\rho,\theta,\Delta}(0).$$

The second moment, denoted by μ_2^+, is

$$\begin{aligned}
\mu_2^+ &= \int_0^\infty x^2 f^+_{\rho,\theta,\Delta}(x)dx = 0^2 \cdot F_{\rho,\theta,\Delta}(0) + \int_0^\infty x^2 f_{\rho,\theta,\Delta}(x)dx = \int_0^\infty x^2 f_{\rho,\theta,\Delta}(x)dx \\
&\overset{(2.3)}{=} \int_0^\infty \Big(\theta^2\rho(\rho+1)f_{\rho+2,\theta,\Delta}(x) - 2\Delta\theta\rho f_{\rho+1,\theta,\Delta}(x) + \Delta^2 f_{\rho,\theta,\Delta}(x)\Big)dx \\
&= \theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) - 2\Delta\theta\rho\big(1-F_{\rho+1,\theta,\Delta}(0)\big) + \Delta^2\big(1-F_{\rho,\theta,\Delta}(0)\big) \\
&= \theta\rho G_{\rho+1,\theta,\Delta}(0) - \Delta G_{\rho,\theta,\Delta}(0).
\end{aligned}$$

We will need the third moment in Section 2.5; the third moment, μ_3^+, is

$$\begin{aligned}
\mu_3^+ &= \int_0^\infty x^3 f^+_{\rho,\theta,\Delta}(x)dx = 0^3 \cdot F_{\rho,\theta,\Delta}(0) + \int_0^\infty x^3 f_{\rho,\theta,\Delta}(x)dx = \int_0^\infty x^3 f_{\rho,\theta,\Delta}(x)dx \\
&\overset{(2.4)}{=} \int_0^\infty \Big(\theta^3\rho(\rho+1)(\rho+2)f_{\rho+3,\theta,\Delta}(x) - 3\Delta\theta^2\rho(\rho+1)f_{\rho+2,\theta,\Delta}(x) \\
&\quad + 3\Delta^2\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta^3 f_{\rho,\theta,\Delta}(x)\Big)dx \\
&= \theta^3\rho(\rho+1)(\rho+2)\big(1-F_{\rho+3,\theta,\Delta}(0)\big) - 3\Delta\theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) \\
&\quad + 3\Delta^2\theta\rho\big(1-F_{\rho+1,\theta,\Delta}(0)\big) - \Delta^3\big(1-F_{\rho,\theta,\Delta}(0)\big) \\
&= \theta^2\rho(\rho+1)G_{\rho+2,\theta,\Delta}(0) - 2\Delta\theta\rho G_{\rho+1,\theta,\Delta}(0) + \Delta^2 G_{\rho,\theta,\Delta}(0).
\end{aligned}$$

The variance of a modified gamma distributed variable with a point mass at zero, denoted by σ^{+^2}, is

$$\begin{aligned}
\sigma^{+^2} &= \mu_2^+ - (\mu^+)^2 = \rho\theta G_{\rho+1,\theta,\Delta}(0) - \Delta G_{\rho,\theta,\Delta}(0) - G^2_{\rho,\theta,\Delta}(0) \\
&= \theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) - 2\Delta\rho\theta\big(1-F_{\rho+1,\theta,\Delta}(0)\big) + \Delta^2\big(1-F_{\rho,\theta,\Delta}(0)\big) \\
&\quad - \Big(\rho^2\theta^2\big(1-F_{\rho+1,\theta,\Delta}(0)\big)^2 \\
&\qquad - 2\rho\theta\Delta\big(1-F_{\rho+1,\theta,\Delta}(0)\big)\big(1-F_{\rho,\theta,\Delta}(0)\big) + \Delta^2\big(1-F_{\rho,\theta,\Delta}(0)\big)^2\Big) \\
&= \theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) + \Delta^2\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big(1-\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big) \\
&\quad - \rho\theta\big(1-F_{\rho+1,\theta,\Delta}(0)\big)\Big(2\Delta + \rho\theta\big(1-F_{\rho+1,\theta,\Delta}(0)\big) - 2\Delta\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big) \\
&= \theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) + \Delta^2 F_{\rho,\theta,\Delta}(0)\big(1-F_{\rho,\theta,\Delta}(0)\big) \\
&\quad - \rho\theta\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big(\rho\theta\big(1-F_{\rho+1,\theta,\Delta}(0)\big) + 2\Delta\Big(1-\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big)\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \theta^2\rho(\rho+1)\big(1-F_{\rho+2,\theta,\Delta}(0)\big) + \Delta^2 F_{\rho,\theta,\Delta}(0)\big(1-F_{\rho,\theta,\Delta}(0)\big) \\
&\quad - \rho\theta\big(1-F_{\rho,\theta,\Delta}(0)\big)\Big(\rho\theta\big(1-F_{\rho+1,\theta,\Delta}(0)\big) + 2\Delta F_{\rho,\theta,\Delta}(0)\Big).
\end{aligned}
$$

Finally, the coefficient of variation, denoted by ν^+, is

$$
\nu^+ = \frac{\sqrt{\rho\theta G_{\rho+1,\theta,\Delta}(0) - \Delta G_{\rho,\theta,\Delta}(0) - G^2_{\rho,\theta,\Delta}(0)}}{G_{\rho,\rho\Delta}(0)}.
$$

2.3.2 Modified gamma distribution truncated at zero

The modified gamma distribution truncated at zero is obtained by ignoring negative realizations of the shifted gamma distribution. The graph of the pdf belonging to the modified gamma distribution that is truncated at zero, is provided in Figure 2.5. The pdf and cdf of this modified distribution, denoted by $f^*_{\rho,\theta,\Delta}$ and $F^*_{\rho,\theta,\Delta}$, are

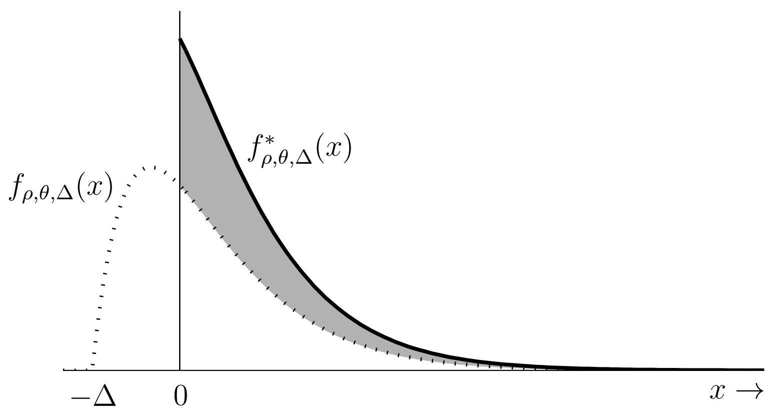


Figure 2.5: The pdf of the modified gamma distribution truncated at zero.

$$
f^*_{\rho,\theta,\Delta}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{f_{\rho,\theta,\Delta}(x)}{1-F_{\rho,\theta,\Delta}(0)} & \text{if } x \geq 0, \end{cases} \tag{2.7}
$$

$$
F^*_{\rho,\theta,\Delta}(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{F_{\rho,\theta,\Delta}(x) - F_{\rho,\theta,\Delta}(0)}{1-F_{\rho,\theta,\Delta}(0)} & \text{if } x \geq 0. \end{cases} \tag{2.8}
$$

Let $\mathcal{P}$ denote $1 - F_{\rho,\theta,\Delta}(0)$. Then the kth moment of the modified shifted gamma distribution truncated at zero, denoted by μ^*_k, is

$$
\mu^*_k = \int_0^\infty x^k f^*_{\rho,\theta,\Delta}(x)dx = \int_0^\infty x^k \frac{f_{\rho,\theta,\Delta}(x)}{\mathcal{P}}dx = \frac{1}{\mathcal{P}}\int_0^\infty x^k f_{\rho,\theta,\Delta}(x)dx = \frac{\mu^+_k}{\mathcal{P}}.
$$

Hence, the expected value, and the second and third moment of the truncated shifted gamma distribution are

$$\mu^* = \frac{\mu^+}{\mathcal{P}}, \quad \mu_2^* = \frac{\mu_2^+}{\mathcal{P}}, \quad \text{and} \quad \mu_3^* = \frac{\mu_3^+}{\mathcal{P}}.$$

The variance of a truncated shifted gamma distribution, denoted by ${\sigma^*}^2$, is

$$\begin{aligned}
{\sigma^*}^2 &= \mu_2^* - (\mu^*)^2 = \frac{\mu_2^+}{\mathcal{P}} - \frac{(\mu^+)^2}{\mathcal{P}^2} = \frac{\big(1 - F_{\rho,\theta,\Delta}(0)\big)\mu_2^+ - (\mu^+)^2}{\mathcal{P}^2} \\
&= \frac{\big(1 - F_{\rho,\theta,\Delta}(0)\big)\Big(\theta^2\rho(\rho+1)\big(1 - F_{\rho+2,\theta,\Delta}(0)\big) + \Delta^2\big(1 - F_{\rho,\theta,\Delta}(0)\big)\Big)}{\mathcal{P}^2} \\
&\quad + \frac{\big(1 - F_{\rho,\theta,\Delta}(0)\big)\Big(-2\Delta\theta\rho\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)\Big) - \rho^2\theta^2\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2} \\
&\quad + \frac{2\rho\theta\Delta\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)\big(1 - F_{\rho,\theta,\Delta}(0)\big) - \Delta^2\big(1 - F_{\rho,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2} \\
&= \frac{\theta^2\rho(\rho+1)\big(1 - F_{\rho,\theta,\Delta}(0)\big)\big(1 - F_{\rho+2,\theta,\Delta}(0)\big) + \Delta^2\big(1 - F_{\rho,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2} \\
&\quad + \frac{-2\Delta\theta\rho\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)\big(1 - F_{\rho,\theta,\Delta}(0)\big) - \rho^2\theta^2\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2} \\
&\quad + \frac{2\rho\theta\Delta\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)\big(1 - F_{\rho,\theta,\Delta}(0)\big) - \Delta^2\big(1 - F_{\rho,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2} \\
&= \frac{\theta^2\rho(\rho+1)\big(1 - F_{\rho,\theta,\Delta}(0)\big)\big(1 - F_{\rho+2,\theta,\Delta}(0)\big) - \rho^2\theta^2\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)^2}{\mathcal{P}^2}.
\end{aligned}$$

The coefficient of variation, ν^*, is

$$\begin{aligned}
\nu^* &= \frac{\theta\sqrt{\rho}\sqrt{(\rho+1)\big(1 - F_{\rho,\theta,\Delta}(0)\big)\big(1 - F_{\rho+2,\theta,\Delta}(0)\big) - \rho\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)^2}}{\mu^+} \\
&= \frac{\theta\sqrt{\rho}\sqrt{(\rho+1)\big(1 - F_{\rho,\theta,\Delta}(0)\big)\big(1 - F_{\rho+2,\theta,\Delta}(0)\big) - \rho\big(1 - F_{\rho+1,\theta,\Delta}(0)\big)^2}}{\rho\theta\big(1 - F_{\rho+1,\theta,\Delta}(0)\big) - \Delta\big(1 - F_{\rho,\theta,\Delta}(0)\big)}.
\end{aligned}$$

2.4 Determination of order-up-to levels

First, we derive the order-up-to levels for the shifted gamma distribution, since this distribution is not widely applied in inventory control. We consider both the P_1 service criterion and the P_2 service criterion.

If the demand, denoted by D, is shifted gamma distributed with demand parameters ρ, θ and Δ, the order-up-to level under the P_1 criterion, denoted by S_α, is found by solving

$$\mathbb{P}\left(D \leq S_\alpha\right) = \alpha.$$

Using that $F_{\rho,\theta,\Delta}(S_\alpha) = \mathbb{P}\left(D \leq S_\alpha\right)$, we obtain

$$S_\alpha = F^{-1}_{\rho,\theta,\Delta}(\alpha).$$

If the fill rate criterion is considered, the order-up-to level, S_β, is determined by solving

$$1 - \frac{\mathbb{E}\left[(D - S_\beta)^+\right]}{\mathbb{E}\left[D\right]} = 1 - \frac{\mathbb{E}\left[(D - S_\beta)^+\right]}{\mu} = \beta.$$

Rewriting the above and using (2.1) leads to the following expression for the order-up-to level:

$$\begin{aligned}(1-\beta)\mu = \mathbb{E}\left[(D - S_\beta)^+\right] &= \int_{S_\beta}^{\infty} (x - S_\beta) f_{\rho,\theta,\Delta}(x)dx = G_{\rho,\theta,\Delta}(S_\beta),\\ S_\beta &= G^{-1}_{\rho,\theta,\Delta}((1-\beta)\mu).\end{aligned}$$

2.4.1 Using $F^{+}_{\rho,\theta,\Delta}$

Let us now assume that demand (D) is distributed according to a modified shifted gamma distribution with a point mass at zero with parameters ρ, θ and Δ. The order-up-to level using the cycle service criterion, denoted by S^+_α is obtained by solving

$$\mathbb{P}\left(D \leq S^+_\alpha\right) = \alpha.$$

The order-up-to level under the P_1 service level is

$$S^+_\alpha = F^{+^{-1}}_{\rho,\theta,\Delta}(\alpha) \overset{(*)}{=} F^{-1}_{\rho,\theta,\Delta}(\alpha) = S_\alpha.$$

Note that the equality at $(*)$ is obtained using (2.6) and that it implicitly assumes that $\alpha > F_{\rho,\theta,\Delta}(0)$. This is not a limitation in practice, since there will not be many SKUs (an SKU is a stock-keeping unit) that have a probability of zero demand that is higher than the desired service level.

If we consider the P_2 service level criterion, the order-up-to level S_β^+ needs to satisfy

$$1 - \frac{\mathbb{E}\left[(D - S_\beta^+)^+\right]}{\mathbb{E}\left[D\right]} = 1 - \frac{\mathbb{E}\left[(D - S_\beta^+)^+\right]}{\mu^+} = \beta.$$

Rewriting and using both (2.5) and (2.1) leads to

$$\begin{aligned}(1-\beta)\mu^+ = \mathbb{E}\left[(D - S_\beta^+)^+\right] &= \int_{S_\beta^+}^{\infty} (x - S_\beta^+) f_{\rho,\theta,\Delta}^+(x) dx \\ &= \int_{S_\beta^+}^{\infty} (x - S_\beta^+) f_{\rho,\theta,\Delta}(x) dx = G_{\rho,\theta,\Delta}(S_\beta^+), \\ S_\beta^+ &= G_{\rho,\theta,\Delta}^{-1}((1-\beta)\mu^+) < S_\beta.\end{aligned}$$

We implicitly assumed that $S_\beta^+ > 0$, which is a reasonable assumption in practice; there will not be many SKUs that have negative order-up-to levels. Furthermore, the inequality is obtained from the fact that $\mu^+ > \mu$ and $G_{\rho,\theta,\Delta}^{-1}(x)$ is a decreasing function, as $G_{\rho,\theta,\Delta}(x)$ is a decreasing function.

2.4.2 Using $F_{\rho,\theta,\Delta}^*$

Let us now assume that demand D is distributed according to a truncated shifted gamma distribution with parameters ρ, θ and Δ. As before, let $\mathcal{P} = 1 - F_{\rho,\theta,\Delta}(0)$ for brevity. The order-up-to level under the P_1 service level, S_α^*, is found by solving

$$\mathbb{P}\left(D \leq S_\alpha^*\right) = \alpha.$$

Rewriting and using (2.8) leads to

$$\begin{aligned}F_{\rho,\theta,\Delta}^*(S_\alpha^*) &= \frac{F_{\rho,\theta,\Delta}(S_\alpha^*) - F_{\rho,\theta,\Delta}(0)}{\mathcal{P}} = \alpha, \\ F_{\rho,\theta,\Delta}(S_\alpha^*) &= \alpha\mathcal{P} + F_{\rho,\theta,\Delta}(0) = \alpha(1 - F_{\rho,\theta,\Delta}(0)) + F_{\rho,\theta,\Delta}(0) \\ &= \alpha + (1-\alpha)F_{\rho,\theta,\Delta}(0), \\ S_\alpha^* &= F_{\rho,\theta,\Delta}^{-1}(\alpha + (1-\alpha)F_{\rho,\theta,\Delta}(0)) > S_\alpha = S_\alpha^+.\end{aligned}$$

The inequality is obtained by noting that $\alpha + (1-\alpha)F_{\rho,\theta,\Delta}(0) > \alpha$ and that $F_{\rho,\theta,\Delta}(x)$ is an increasing function, and therefore also $F_{\rho,\theta,\Delta}^{-1}(x)$ is. Again, we need that the order-up-to level is positive.

If a fill rate criterion is used, the order-up-to level, denoted by S^*_β, is determined using

$$1 - \frac{\mathbb{E}\left[(D - S^*_\beta)^+\right]}{\mathbb{E}\left[D\right]} = 1 - \frac{\mathbb{E}\left[(D - S^*_\beta)^+\right]}{\mu^*} = \beta.$$

Rewriting the above and using (2.7) provides an expression for the order-up-to level, namely

$$\begin{aligned}(1-\beta)\mu^* = \mathbb{E}\left[(D - S^+_\beta)^*\right] &= \int_{S^*_\beta}^{\infty} (x - S^*_\beta) f^*_{\rho,\theta,\Delta}(x)dx \\ &= \int_{S^*_\beta}^{\infty} (x - S^*_\beta) \frac{f_{\rho,\theta,\Delta}(x)}{\mathcal{P}} dx = \frac{1}{\mathcal{P}} G_{\rho,\theta,\Delta}(S^*_\beta), \\ S^*_\beta = G^{-1}_{\rho,\theta,\Delta}((1-\beta)\mu^*\mathcal{P}) &\overset{(*)}{=} G^{-1}_{\rho,\theta,\Delta}((1-\beta)\mu^+) = S^+_\beta.\end{aligned}$$

Also here we need that $S^*_\beta > 0$. For the equality at $(*)$ we use that $\mu^* = \frac{\mu^+}{\mathcal{P}}$. Note that since $S^+_\beta < S_\beta$ it also holds that $S^*_\beta < S_\beta$.

2.4.3 Results with general distributions

Note that for both the modified normal distribution and the modified shifted gamma distribution the following properties for the order-up-to levels hold (see Sections 2.2 and 2.4):

$$\begin{array}{lcccl} S^+_\alpha & = & S_\alpha & \leq & S^*_\alpha, \\ S^*_\beta & = & S^+_\beta & \leq & S_\beta. \end{array} \tag{2.9}$$

In the remainder of this section it is proven to be true for any modified distribution constructed in the same way.

Let $f(x)$ and $F(x)$ be a general density and distribution function of a continuous distribution. Now let us construct the two modified distributions $f^+(x)$ and $F^+(x)$, and $f^*(x)$ and $F^*(x)$; the first has a point mass at zero with probability $F(0)$ and the second is truncated at zero. The pdf and cdf of the distribution with a point mass are defined as follows:

$$f^+(x) = \begin{cases} 0 & \text{if } x < 0 \\ F(0) & \text{if } x = 0 \\ f(x) & \text{if } x > 0, \end{cases} \qquad \text{and} \qquad F^+(x) = \begin{cases} 0 & \text{if } x < 0 \\ F(x) & \text{if } x \geq 0. \end{cases}$$

The pdf and cdf of the truncated distribution are

$$f^*(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{f(x)}{1-F(0)} & \text{if } x \geq 0, \end{cases} \qquad \text{and} \qquad F^*(x) = \begin{cases} 0 & \text{if } x < 0 \\ \dfrac{F(x)-F(0)}{1-F(0)} & \text{if } x \geq 0. \end{cases}$$

Note that in case only nonnegative realizations are possible under the original distribution (so, if $F(0) = 0$), the original distribution and the modified distributions coincide. In that case also the order-up-to levels coincide, hence the conjecture that (2.9) will hold in general is trivial. Therefore, for the remainder of this section we assume that $F(0) > 0$ and in that case strict inequalities will hold in (2.9).

Further, the expected value of the distribution with a point mass, denoted by μ^+ is given by

$$\mu^+ = \int_{-\infty}^{\infty} x f^+(x) dx = 0F(0) + \int_0^{\infty} x f(x) dx = \int_0^{\infty} x f(x) dx,$$

while the expected value of the truncated distribution, μ^*, is

$$\mu^* = \int_{-\infty}^{\infty} x f^*(x) dx = \int_0^{\infty} x \frac{f(x)}{1 - F(0)} dx = \frac{1}{1 - F(0)} \int_0^{\infty} x f(x) dx.$$

Hence, $\mu^* = \frac{\mu^+}{1-F(0)}$.

Let $G(z)$ denote the loss function belonging to a variable X that is distributed according to $f(x)$, so

$$G(z) = \int_z^{\infty} (x - z) f(x) dx = \mathbb{E}\left[(X - z)^+\right].$$

The order-up-to levels of the original distribution under the P_1 and P_2 service equation, denoted by S_α and S_β respectively, are

$$S_\alpha = F^{-1}(\alpha),$$
$$S_\beta = G^{-1}((1 - \beta)\mu),$$

where μ denotes the expected value of the original distribution. Note that

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{0} x f(x) dx + \int_0^{\infty} x f(x) dx < \int_0^{\infty} x f(x) dx = \mu^+.$$

Now let us consider the case that demand (D) is distributed according to the distribution with a point mass at zero. The order-up-to levels under the cycle service and fill rate criterion are denoted by S_α^+ and S_β^+ respectively. First consider the P_1 service criterion. The order-up-to level is chosen such that

$$\alpha = F^+(S_\alpha^+) = F(S_\alpha^+).$$

The order-up-to level is

$$S_\alpha^+ = F^{-1}(\alpha) = S_\alpha,$$

so indeed $S_\alpha^+ = S_\alpha$. For the fill rate service criterion the order-up-to level needs to satisfy

$$1 - \frac{\mathbb{E}\left[(D - S_\beta^+)^+\right]}{\mathbb{E}\left[D\right]} = 1 - \frac{\mathbb{E}\left[(D - S_\beta^+)^+\right]}{\mu^+} = \beta.$$

The order-up-to level is obtained by rewriting:

$$\begin{aligned}(1-\beta)\mu^+ &= \int_{S_\beta^+}^\infty (x - S_\beta^+) f^+(x) dx \\ &= \int_{S_\beta^+}^\infty (x - S_\beta^+) f(x) dx = G(S_\beta^+), \\ S_\beta^+ &= G^{-1}((1-\beta)\mu^+).\end{aligned}$$

Since $\mu^+ > \mu$ and $G(x)$ is a strictly decreasing function, we know that $G^{-1}((1-\beta)\mu) > G^{-1}((1-\beta)\mu^+)$ and hence

$$S_\beta^+ < S_\beta.$$

Finally, let us consider the case that demand D is distributed according to the truncated distribution. The order-up-to levels under the cycle service and fill rate criterion are S_α^* and S_β^* respectively. First, S_α^* is chosen such that $F^*(S_\alpha^*) = \alpha$. Rewriting this leads to

$$\begin{aligned}F^*(S_\alpha^*) &= \frac{F(S_\alpha^*) - F(0)}{1 - F(0)} = \alpha, \\ F(S_\alpha^*) &= \alpha(1 - F(0)) + F(0) = \alpha + (1-\alpha)F(0), \\ S_\alpha^* &= F^{-1}(\alpha + (1-\alpha)F(0)).\end{aligned}$$

Since $\alpha + (1-\alpha)F(0) > \alpha$ and $F(x)$ is a strictly increasing function, we know that $F^{-1}(\alpha + (1-\alpha)F(0)) > F^{-1}(\alpha)$ and hence

$$S_\alpha^* > S_\alpha.$$

Consider S_β^*. This order-up-to level needs to satisfy $1 - \frac{\mathbb{E}\left[(D - S_\beta^*)^+\right]}{\mathbb{E}[D]} = \beta$; rewriting provides us with

$$\begin{aligned}(1-\beta)\mu^* &= \int_{S_\beta^*}^\infty (x - S_\beta^*) f^*(x) dx = \int_{S_\beta^*}^\infty (x - S_\beta^*) \frac{f(x)}{1 - F(0)} dx = \frac{1}{1 - F(0)} G(S_\beta^*), \\ S_\beta^* &= G^{-1}((1-\beta)\mu^*(1 - F(0))).\end{aligned}$$

Since $\mu^* = \frac{\mu^+}{1-F(0)}$, we know that $G^{-1}((1-\beta)\mu^*(1-F(0))) = G^{-1}((1-\beta)\mu^+)$, hence,

$$S^*_\beta = S^+_\beta.$$

Note that for rewriting the modified distributions to the original distribution we need that the order-up-to levels are positive. This is not a limitation in real life, since negative order-up-to levels will not appear in practice.

2.5 Mistakenly using the (shifted) gamma distribution

In Section 2.4 we have derived the order-up-to levels under the assumption that demand is distributed according to two modified shifted gamma distributions. In this section we consider using the regular and the shifted gamma distribution while in fact the demand is modified shifted gamma distributed. A modified shifted gamma distribution and a regular gamma distribution may appear to be quite similar, hence one could mistakenly choose to use a regular gamma distribution instead of one of the two modified shifted gamma distributions. In the remainder of this section we determine what happens to the attained service level when this occurs. One will probably not choose to fit a shifted gamma distribution, because of the occurrence of negative realizations. However, for the sake of completeness the use of the shifted gamma distribution while one should use the modified shifted gamma distribution, is also discussed in this section.

The order-up-to levels using the shifted gamma distribution are provided at the start of Section 2.4; note that the parameters as a function of the first three moments of the shifted gamma distribution are (see Appendix A.2):

$$\theta = \frac{\mu_3 + 2\mu^3 - 3\mu\mu_2}{2(\mu_2 - \mu^2)}, \tag{2.10}$$

$$\rho = \frac{\mu_2 - \mu^2}{\theta^2} = \frac{\sigma^2}{\theta^2}, \tag{2.11}$$

$$\Delta = \frac{\mu_2 - \mu^2}{\theta} - \mu = \frac{\sigma^2}{\theta} - \mu. \tag{2.12}$$

The order-up-to levels using a regular gamma distribution, denoted by S^o_α (cycle service) and S^o_β (fill rate), are:

$$S^o_\alpha = F^{-1}_{\rho,\theta}(\alpha),$$
$$S^o_\beta = G^{-1}_{\rho,\theta}((1-\beta)\mu),$$

where $F_{\rho,\theta} = F_{\rho,\theta,0}$ and $G_{\rho,\theta} = G_{\rho,\theta,0}$. The parameters as a function of the first two moments are

$$\theta = \frac{\mu_2 - \mu^2}{\mu} = \frac{\sigma^2}{\mu}, \tag{2.13}$$

$$\rho = \frac{\mu^2}{\mu_2 - \mu^2} = \frac{\mu^2}{\sigma^2}. \tag{2.14}$$

We first consider that demand is distributed according to the modified shifted gamma distribution with a point mass at zero, next the truncated shifted gamma distribution is considered.

2.5.1 Not using $F^+_{\rho,\theta,\Delta}$

If we know that demand is distributed according the modified shifted gamma distribution with a point mass at zero with demand parameters ρ, θ and Δ, the first three moments are (see Section 2.3.1)

$$\begin{aligned}
\mu^+ &= G_{\rho,\theta,\Delta}(0),\\
\mu_2^+ &= \theta\rho G_{\rho+1,\theta,\Delta}(0) - \Delta G_{\rho,\theta,\Delta}(0),\\
\mu_3^+ &= \theta^2\rho(\rho+1)G_{\rho+2,\theta,\Delta}(0) - 2\Delta\theta\rho G_{\rho+1,\theta,\Delta}(0) + \Delta^2 G_{\rho,\theta,\Delta}(0).
\end{aligned}$$

We can estimate the parameters of the shifted and regular gamma distribution by substituting these three moments in (2.10)–(2.14). Let us denote these parameters by $\hat{\theta}^+$, $\hat{\rho}^+$ and $\hat{\Delta}^+$ for the shifted gamma distribution, and $\tilde{\theta}^+$ and $\tilde{\rho}^+$ for the regular gamma distribution, so, e.g.,

$$\hat{\theta}^+ = \frac{\mu_3^+ + 2{\mu^+}^3 - 3\mu^+\mu_2^+}{2(\mu_2^+ - {\mu^+}^2)}.$$

This leads to the estimated order-up-to levels

$$\begin{aligned}
\widehat{S}_\alpha^+ &= F^{-1}_{\hat{\rho}^+,\hat{\theta}^+,\hat{\Delta}^+}(\alpha),\\
\widehat{S}_\beta^+ &= G^{-1}_{\hat{\rho}^+,\hat{\theta}^+,\hat{\Delta}^+}((1-\beta)\mu^+),\\
\widetilde{S}_\alpha^+ &= F^{-1}_{\tilde{\rho}^+,\tilde{\theta}^+}(\alpha),\\
\widetilde{S}_\beta^+ &= G^{-1}_{\tilde{\rho}^+,\tilde{\theta}^+}((1-\beta)\mu^+).
\end{aligned}$$

Since the true demand distribution is known, we can also obtain the attained service levels using the wrong order-up-to levels. The attained service using the shifted gamma distribution, denoted by $\widehat{\alpha}^+$ and $\widehat{\beta}^+$ for the P_1 and P_2 service level respectively, is

$$\widehat{\alpha}^+ = F^+_{\rho,\theta,\Delta}(\widehat{S}^+_\alpha),$$
$$\widehat{\beta}^+ = 1 - \frac{G_{\rho,\theta,\Delta}(\widehat{S}^+_\beta)}{\mu^+}.$$

Finally, the attained service levels using the regular gamma distribution, $\widetilde{\alpha}^+$ and $\widetilde{\beta}^+$, are

$$\widetilde{\alpha}^+ = F^+_{\rho,\theta,\Delta}(\widetilde{S}^+_\alpha),$$
$$\widetilde{\beta}^+ = 1 - \frac{G_{\rho,\theta,\Delta}(\widetilde{S}^+_\beta)}{\mu^+}.$$

Figures 2.6–2.9 show the attained service level when using $\widehat{S}^+_\alpha$, $\widehat{S}^+_\beta$, $\widetilde{S}^+_\alpha$ and $\widetilde{S}^+_\beta$, respectively, for two values of the desired service level, for three values of ρ, and for different values of Δ. The value of θ is $\frac{10}{\rho}$ for all graphs. Also other values of θ are used and these result in similar graphs (not shown here). Figure 2.6 shows the

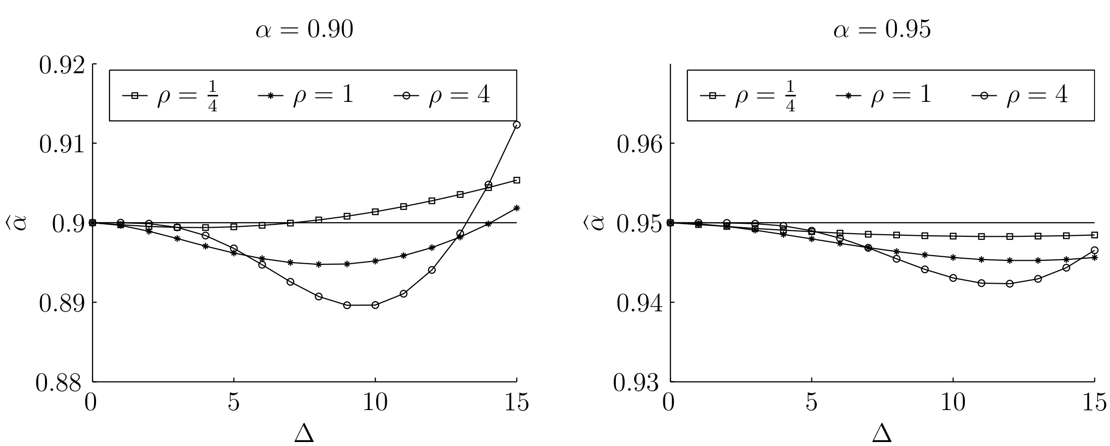


Figure 2.6: Attained P_1 service when using a shifted gamma distribution, while demand is distributed according to $F^+_{\rho,\theta,\Delta}$.

attained cycle service when using the shifted gamma distribution, while in fact the demand is modified shifted gamma distributed with a point mass at zero. In both the left graph ($\alpha = 0.90$) and the right graph ($\alpha = 0.95$) the attained service level is close to the desired service level. In case of $\alpha = 0.95$ the attained service is below the desired service for all values of Δ, but the largest underperformance is less than

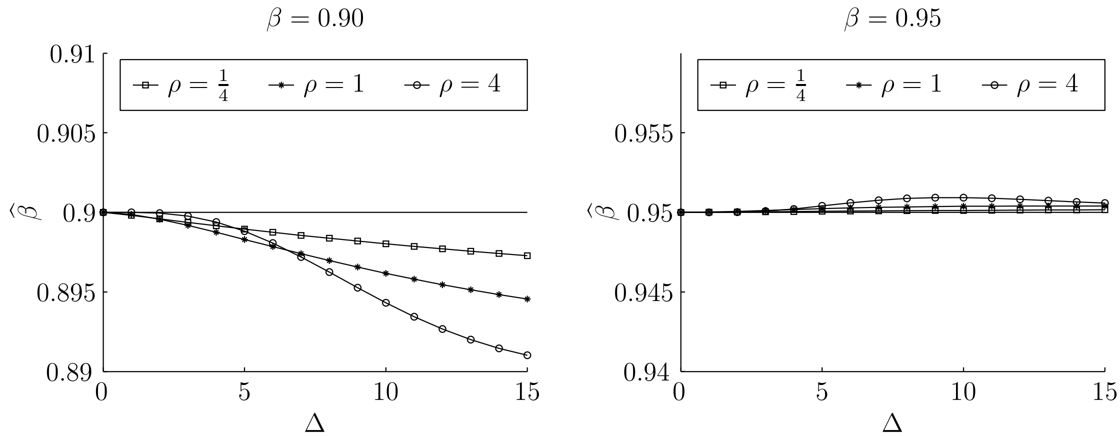


Figure 2.7: Attained P_2 service when using a shifted gamma distribution, while demand is distributed according to $F^+_{\rho,\theta,\Delta}$.

0.01. In case of $\alpha = 0.90$, the largest underperformance is close to 0.01, while also overperformance exist, which is at most 0.01, unless Δ and ρ are large.

Figure 2.7 shows the attained fill rate when assuming a shifted gamma distribution while demand is actually distributed according to a modified gamma distribution with a point mass at zero. In case of $\beta = 0.90$ we see underperformance for all values of Δ and the larger Δ, the larger the underperformance. However, the largest underperformance depicted is less then 0.01, so not very large. In case of $\beta = 0.95$ the attained service is very close to the desired service level, so using the shifted gamma distribution does not result in very different order-up-to levels.

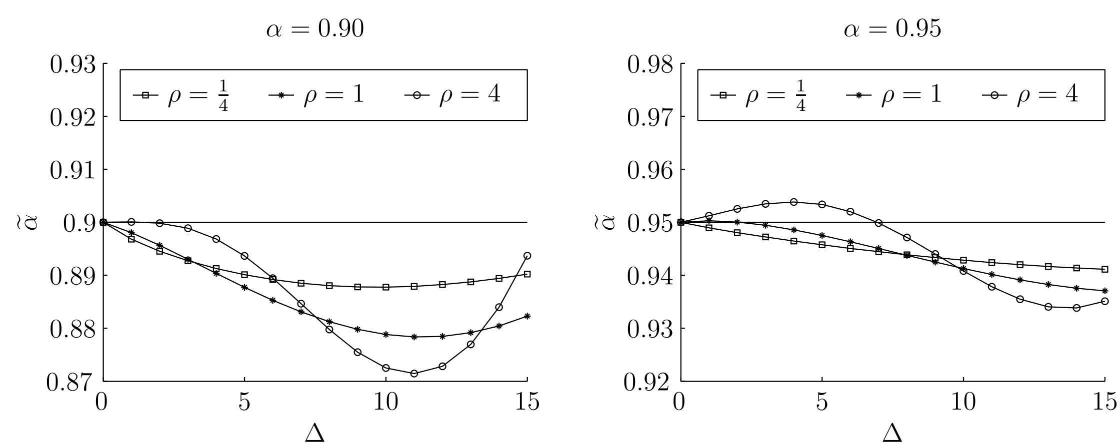


Figure 2.8: Attained P_1 service when using a regular gamma distribution, while demand is distributed according to $F^+_{\rho,\theta,\Delta}$.

In Figure 2.8, showing the attained P_1 service when using a regular gamma distribution, we see that when $\alpha = 0.90$, the desired service is only reached for small values of Δ; otherwise there exist underperformance, which is at most almost 0.03. The underperformance first is getting bigger when Δ is getting bigger, but it is decreasing after a certain point. In case of $\alpha = 0.95$ the attained service levels show approximately the same pattern; they are a little closer to the desired service levels and in case of $\rho = 4$, we even see a small overperformance for relative small values for Δ. The largest underperformance is almost 0.02 in this case.

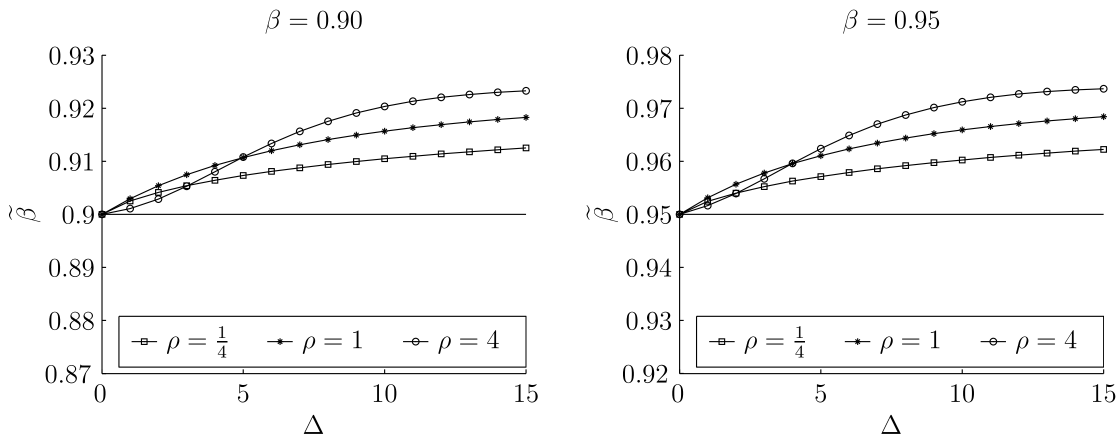


Figure 2.9: Attained P_2 service when using a regular gamma distribution, while demand is distributed according to $F^+_{\rho,\theta,\Delta}$.

Figure 2.9, finally, shows the attained fill rates when using a regular gamma distribution while we should use the modified gamma distribution. The graphs for $\beta = 0.90$ and $\beta = 0.95$ are quite similar. Both show overperformance, being at most a little over 0.02.

In general the under- and overperformance is not large. Therefore, it seems not to lead to big problems when using the wrong distribution. However, also a small underperformance could lead to loss of clientele if it happens frequently. Overperformance, on the other side, means that too much inventory is kept. Even a relative small reduction in inventory could lead to large savings, if we consider a company with large amounts of SKUs. Furthermore, the attained service level is closer to the desired one if we consider a shifted gamma distribution compared to a regular gamma distribution. This can be explained by the fact that the modified shifted gamma distributions are closer to the shifted gamma distribution than to the regular gamma distribution.

2.5.2 Not using $F^*_{\rho,\theta,\Delta}$

If we know that demand is distributed according to the shifted gamma distribution truncated at zero with parameters ρ, θ and Δ, the first three moments are (see Section 2.3.2)

$$\mu^* = \frac{\mu^+}{\mathcal{P}}, \qquad \mu_2^* = \frac{\mu_2^+}{\mathcal{P}}, \qquad \text{and} \qquad \mu_3^* = \frac{\mu_3^+}{\mathcal{P}},$$

where $\mathcal{P} = 1 - F_{\rho,\theta,\Delta}(0)$. Using these moments and the functions (2.10)–(2.12), we can obtain estimates for the parameters of the shifted gamma distribution, denoted by $\hat{\rho}^*$, $\hat{\theta}^*$ and $\hat{\Delta}^*$. The order-up-to levels using the shifted gamma distribution, while the true distribution is the truncated shifted gamma distribution, are

$$\widehat{S}^*_\alpha = F^{-1}_{\hat{\rho}^*,\hat{\theta}^*,\hat{\Delta}^*}(\alpha),$$
$$\widehat{S}^*_\beta = G^{-1}_{\hat{\rho}^*,\hat{\theta}^*,\hat{\Delta}^*}((1-\beta)\mu^*),$$

for the cycle service and the fill rate service respectively. The attained service, $\widehat{\alpha}^*$ and $\widehat{\beta}^*$ is determined using that demand is truly truncated shifted gamma distributed:

$$\widehat{\alpha}^* = F^*_{\rho,\theta,\Delta}(\widehat{S}^*_\alpha),$$
$$\widehat{\beta}^* = 1 - \frac{\frac{1}{\mathcal{P}} G_{\rho,\theta,\Delta}(\widehat{S}^*_\beta)}{\mu^*}.$$

Figures 2.10 and 2.11 show the attained service when a shifted gamma distribution is used, while the demand is actually truncated shifted gamma distributed for different values of ρ and Δ. The parameter θ is chosen according to $\frac{10}{\rho}$. Figure 2.10 shows the attained P_1 service for $\alpha = 0.90$ and $\alpha = 0.95$. The left and right graph are quite similar, with under- and overperformance both occurring and both being small, i.e., less then 0.005. Note that in case of $\rho = 1$ the desired service level is reached exactly for all values of Δ.

Figure 2.11 shows the attained fill rates. The graphs are actually quite similar to the graphs in Figure 2.10; the main difference is that the attained service is even closer to the desired service in case of $\beta = 0.95$. We can see that also in these graphs the desired service is exactly reached in case of $\rho = 1$.

Note that the over- and underperformance is not bigger than 0.005, hence using the shifted gamma distribution while in fact we should be using the truncated shifted gamma distribution does not result in big problems.

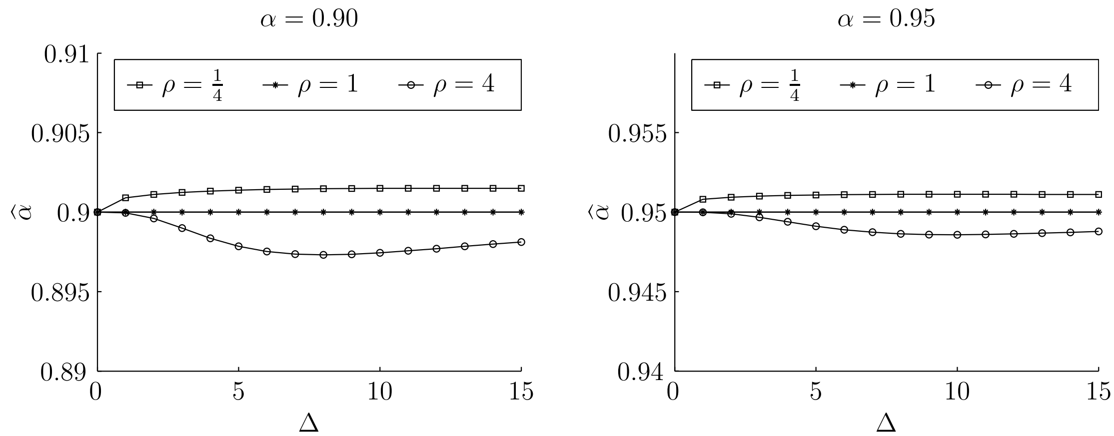


Figure 2.10: Attained P_1 service when using a shifted gamma distribution, while demand is distributed according to $F^*_{\rho,\theta,\Delta}$.

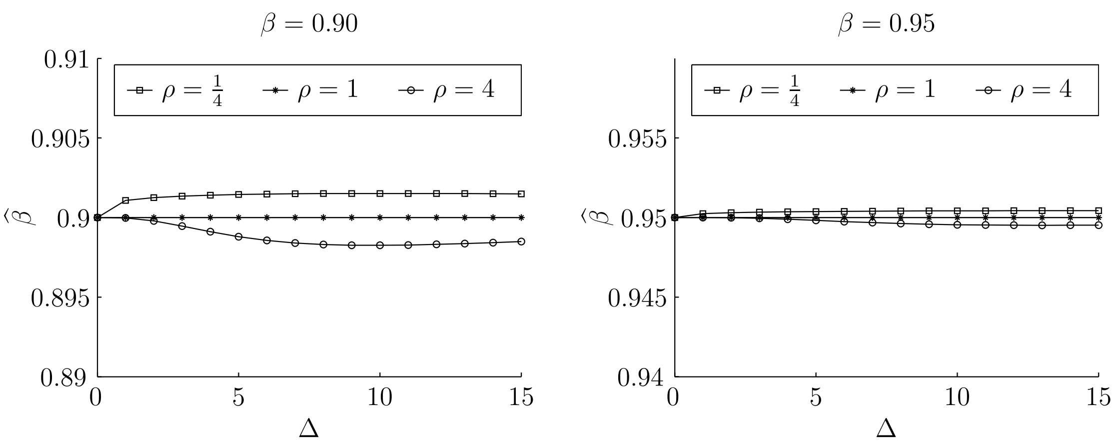


Figure 2.11: Attained P_2 service when using a shifted gamma distribution, while demand is distributed according to $F^*_{\rho,\theta,\Delta}$.

Estimates for the regular gamma distribution, denoted by $\tilde{\rho}^*$ and $\tilde{\theta}^*$, are obtained using μ^*, μ_2^* and (2.13) and (2.14). The order-up-to levels for the P_1 and P_2 service level are, respectively,

$$\widetilde{S}^*_\alpha = F^{-1}_{\tilde{\rho}^*,\tilde{\theta}^*}(\alpha),$$
$$\widetilde{S}^*_\beta = G^{-1}_{\tilde{\rho}^*,\tilde{\theta}^*}((1-\beta)\mu^*).$$

The attained service levels, using that demand is distributed according to the truncated shifted gamma distribution, are

$$\widetilde{\alpha}^* = F^*_{\rho,\theta,\Delta}(\widetilde{S}^*_\alpha),$$
$$\widetilde{\beta}^* = 1 - \frac{\frac{1}{P} G_{\rho,\theta,\Delta}(\widetilde{S}^*_\beta)}{\mu^*}.$$

Figures 2.12 and 2.13 show the attained service when the regular gamma distribution is used while demand is distributed according to the truncated modified shifted gamma distribution for different values of ρ and Δ; θ is determined according to $\frac{10}{\rho}$. Note that also in these figures the desired service level is exactly attained in case of $\rho = 1$. Figure 2.12 show the attained cycle service for $\alpha = 0.90$ and $\alpha = 0.95$. The

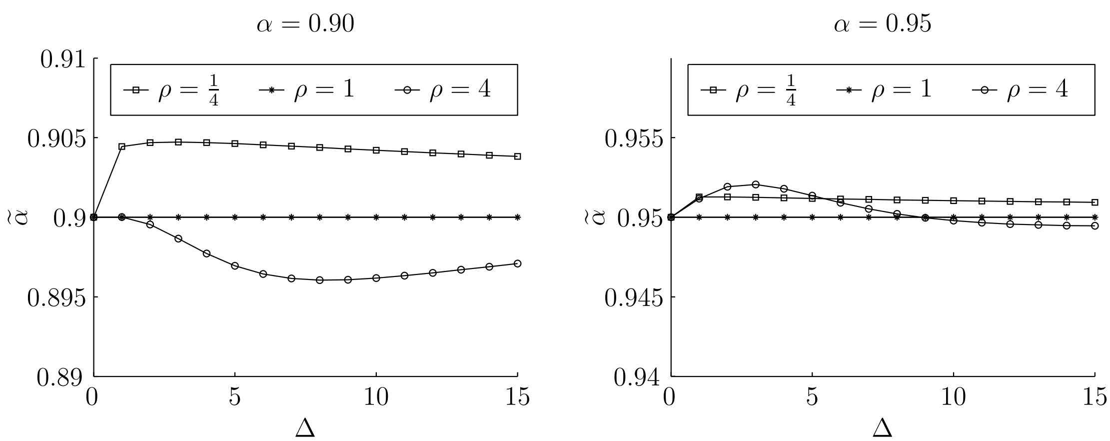


Figure 2.12: Attained P_1 service when using a regular gamma distribution, while demand is distributed according to $F^*_{\rho,\theta,\Delta}$.

left and right graph show both under- and overperformance, but the deviations from the desired service level are small: all attained service levels are within 0.005 from the desired service. Note that for $\alpha = 0.90$ the deviations are a little larger compared to the deviations when using the shifted gamma distribution, which is logical, since the shifted gamma distribution is closer to the truncated shifted gamma distribution compared to the regular gamma distribution.

Figure 2.13 show the attained P_2 service for $\beta = 0.90$ and $\beta = 0.95$. The left and right graph are quite similar: both show small underperformance in case of $\rho = \frac{1}{4}$ and overperformance in case of $\rho = 4$. The deviations from the desired service level are small: all attained service levels are within 0.005 from the desired service. The deviations for both $\beta = 0.90$ and $\beta = 0.95$ are a little larger compared to the deviations in 2.11; this is intuitively explained by the fact that the shifted gamma

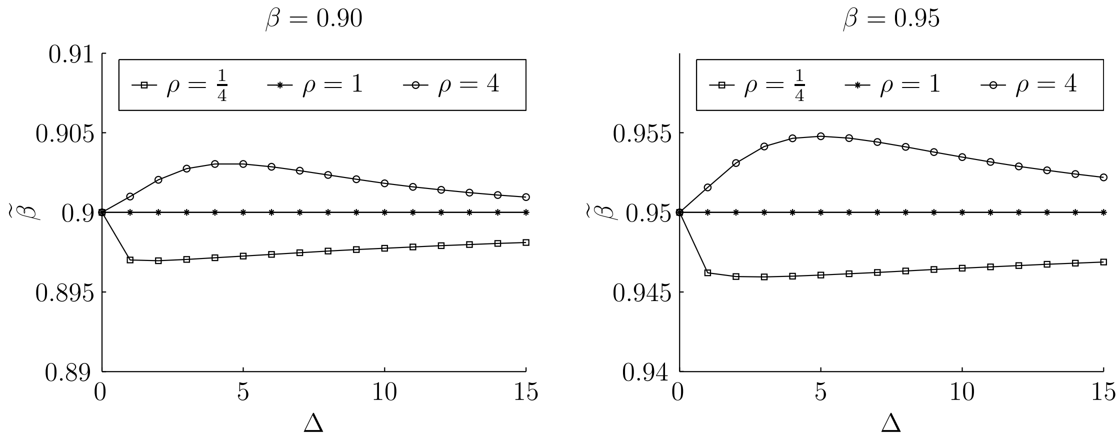


Figure 2.13: Attained P_2 service when using a regular gamma distribution, while demand is distributed according to $F^*_{\rho,\theta,\Delta}$.

distribution is more similar to the truncated shifted gamma distribution than the regular gamma distribution.

2.6 Summary results

This chapter discusses two modified shifted gamma distributions: one that has a point mass at zero and one that is truncated at zero. These distributions provide more flexible ways to model demand, at least when the lead time is zero, including the possibility to model intermittent demand using one distribution instead of using a compound distribution. The regular gamma distribution is a special case of both modified shifted gamma distributions.

These two modified gamma distribution are used to model demand in a periodic review, order-up-to level inventory control policy with zero lead time under two service level restrictions. Using these distributions leads to nice expressions for the order-up-to levels, hence, it is not very difficult to implement these distributions in an inventory control system. Furthermore, we find that the order-up-to levels of the shifted gamma distribution and the modified shifted gamma distribution with a point mass at zero are equal when considering the cycle service. Besides that, the order-up-to levels of the two modified shifted gamma distributions coincide when using the fill rate service. Strijbosch and Moors (2006) found the same results for the normal distribution and Section 2.4.3 shows that this holds for any modified distribution with a point mass at zero and truncated modified distribution if they are constructed in the same way as in this chapter.

Finally we show that using either the gamma distribution or the shifted gamma distribution when demand is actually distributed according to one of the two modified shifted gamma distributions leads to not reaching the desired service level (unless we consider the truncated shifted gamma distribution and $\rho = 1$). The figures in Section 2.5 show that both under- and overperformance occur and that the attained service is close to the desired one: the largest deviation from the desired service level is 0.03. However, even this small underperformance could have negative effects. If one would have negotiated a certain service and in the long run this service is not reached, customers could choose another supplier. Also, overperformance has its disadvantages, since this implies that too much inventory is kept. Less inventory would results in less inventory costs and even a relative little decrease could already lead to large cost savings, as in many companies inventory costs are a substantial part of the costs.

If these distributions are used in a real life inventory control policy, we need to estimate the three parameters ρ, θ and Δ. We could use the expressions of the first three moments of the modified demand distributions to estimate the demand parameters. Of course, the true values of the first three moments are not known, hence, we need to substitute the first, second and third moment by estimates of these three moments. These estimates are obtained using historical demand observations. Another possibility for finding estimates of the three demand parameters, is using maximum likelihood estimators. Note that using estimators for the demand parameters will lead to randomness in the order-up-to level, as estimators themselves are random variables and the order-up-to level is a function of the estimators. This will probably lead to underperformance, that can be serious, as is shown in Chapters 3–5 for normally distributed demand, gamma distributed demand, and mixed Erlang distributed demand, respectively.

Part II

Unknown demand parameters

Chapter 3

Normal demand with unknown demand parameters

Inventory models need information about the demand distribution. In practice, this information is not known with certainty and has to be estimated with often relatively few historical demand observations. Using these estimates leads to underperformance. This chapter focuses on normal distributed demand and an (R, S) inventory control policy, where the order-up-to level satisfies a service equation. The cause and size of this underperformance are determined. Furthermore the formulae for the order-up-to levels are corrected analytically where possible and otherwise by use of simulation and linear regression. Simulation shows that these corrections improve the attained service. This chapter is based on Janssen et al. (2009).

3.1 Introduction

This chapter considers the assumption that demand is normally distributed; the normality assumption is often used both in research and in practice, see Zeng and Hayya (1999). Furthermore, see Strijbosch and Moors (2006) for references to recent articles that involve the normal distribution. The assumption of normality is made because it yields tractable results and it seems to give quite good approximations when used on demand data with a low coefficient of variation (Silver et al., 1998 and Zipkin, 2000). However, the normal distribution has two major disadvantages: it is symmetric and can take on negative values, while demand is nonnegative and often skewed to the right. This may not impose serious problems if the coefficient of variation is low (Zipkin, 2000) or if the demand during review/lead time consists of many individual and independent demands (Silver et al., 1998). However, for high values of the coefficient of variation, these disadvantages get more important; hence Strijbosch and Moors (2006) suggest two simple modifications of the normal distribution (see also Section 2.2). Tyworth and O'Neill (1997) and Lau and Lau (2003) investigate the (non)robustness of using the normal approximation with a reorder-point, order quantity inventory policy.

We assume that demand is truly normally distributed, but that the parameters of the normal distribution (the mean and variance) are unknown. These parameters are estimated using historical demand observations and the effect of estimating the parameters is studied. Silver and Rahnama (1986, 1987) have considered the effect of estimating parameters when using a reorder-point, order quantity inventory policy with a cost criterion. They construct a function that determines the expected cost of estimating the demand distribution rather than knowing it, and they conclude that this function is not symmetrical: underestimating the demand causes larger costs than overestimating. In the second article they propose a method that deliberately biases the reorder point upwards.

This chapter focuses on two service level criteria within an (R, S) inventory policy with zero lead time. The independent and identically distributed demands during review periods have a normal distribution with mean μ and standard deviation σ, which leads to a coefficient of variation $\nu = \sigma/\mu$. As mentioned before, the normal distribution could lead to negative demand and in our model this is interpreted as returned goods, so demand is actually net demand (demanded goods minus returned goods). In addition the goods returned by customers can be sent back to the supplier, thus the inventory level at the start of a review period always equals S. Furthermore, demand during $t+1$ consecutive review periods is assumed to be stationary, and the first t periodic review demands are used to estimate the mean and standard deviation of the demand in review period $t+1$. The mean and standard deviation are estimated by their sample statistics. We prefer using sample statistics instead of exponential smoothing, since derivations are more tractable, while the conclusions are similar. In forecasting, using sample statistics in order to obtain estimates is often referred to as (simple) moving average.

Section 3.2 discusses the P_1 service criterion in short. An analytical correction of the order-up-to level is given for the case that only μ is unknown. Section 3.3 focuses on the P_2 service criterion. First two theoretical situations are considered for illustrative purposes: the case that μ is unknown, but σ and ν are known, and the case that μ and σ are unknown, but ν is known. The main Sections 3.3.3 and 3.3.4 treat the important case that these three parameters are all unknown. We show by simulation that just plugging in estimates leads to serious underperformance. Besides, we develop a correction function for the safety factor that nearly gives the desired fill rates. The last section provides a short summary of the results.

3.2 P_1 service level criterion

This section considers the P_1 service criterion. It is common to express the order-up-to level as a function of the mean μ, the standard deviation σ, and a safety factor c_α. Since demand is normally distributed, the order-up-to level is (see, e.g., Silver et al., 1998)

$$S(\mu, \sigma, c_\alpha) = \mu + c_\alpha \sigma. \tag{3.1}$$

The safety factor c_α is $\Phi^{-1}(\alpha)$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution. S without arguments is used to denote the theoretically correct order-up-to level when all parameters are known, so $S = S(\mu, \sigma, c_\alpha)$ in case of a P_1 service criterion.

In practice, the mean and variance are unknown, which means that S is unknown too. The common solution is to replace the parameters μ and σ by its estimates. If we (unrealistically) assume that only μ is unknown and use the sample mean $\bar{d}_t$ to estimate it, the resulting order-up-to level is $S(\bar{d}_t, \sigma, c_\alpha)$ with $\bar{d}_t = \frac{1}{t}\sum_{i=1}^{t} d_i$, where d_i is the observed demand during period i and t is the number of historical demand observations. This order-up-to level, although unbiased, does not meet the service requirements in the long run. Consider Figure 3.1. Since $S(\bar{d}_t, \sigma, c_\alpha)$ is normally distributed, with mean S and variance ${\sigma^2}/{t}$, it is symmetric and the realization $S - \varepsilon$ is equally likely as the realization $S + \varepsilon$. The realization $S + \varepsilon$ decreases the probability of having backlogged demand with the darker area in Figure 3.1, while the realization $S - \varepsilon$ increases the probability of having backlogged demand with the lighter area. The surface of the lighter area is larger than the surface of the darker area and this implies that in the long run the achieved service level falls short of the desired one. This phenomenon is mathematically explained by (τ denotes $\sqrt{1 + 1/t}$)

$$\mathbb{P}\left(D_R < S(\bar{d}_t, \sigma, c_\alpha)\right) = \mathbb{P}\left(D_R - \bar{d}_t < c_\alpha \sigma\right) = \Phi\left(\frac{c_\alpha}{\tau}\right) < \Phi\left(c_\alpha\right) = \alpha, \tag{3.2}$$

if $c_\alpha > 0$, hence $\alpha > 0.5$. Note that $\alpha > 0.50$ is not a limitation in practice: in real life the desired service level will be larger than 0.50.

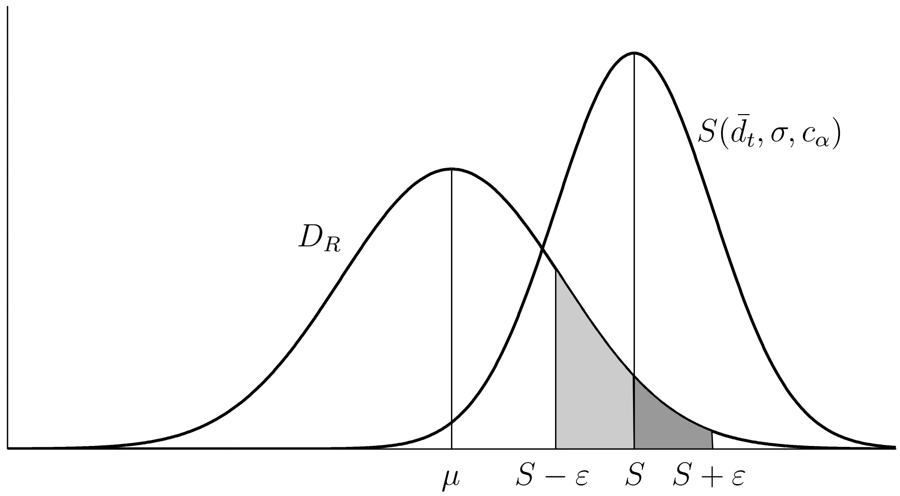


Figure 3.1: The PDF of demand during review D_R and order-up-to level $S(\bar{d}_t, \sigma, c_\alpha)$.

Now let Z be a general biased estimator of S and let $Z \sim N(S + v\sigma, (w^2-1)\sigma^2)$, where $w \geq 1$. The probability of not having backorders in a review period is given by

$$\mathbb{P}\left(D_R < Z\right) = \Phi\left(\frac{c_\alpha + v}{w}\right). \tag{3.3}$$

In general, this expression does not equal α, unless v and w are chosen according to the relation $v = (w-1)c_\alpha$ (see also Strijbosch et al., 1997). This phenomenon is depicted in Figure 3.2. The curved surface depicts the attained service level at values of $0 \leq v \leq 1.6$ and $1 \leq w \leq 1.5$. The fine grid depicts the desired service level. The attained service is below this level in some cases, while it is above it in other cases. The attained service reaches the required one only on the solid line, for which obviously holds $v = (w-1)c_\alpha$. Furthermore, if an unbiased estimator is used ($v = 0$, the dashed line), the desired service is reached only if $w = 1$. That case corresponds to having an estimator with standard deviation equal to 0, which is not possible in practice. So, an unbiased normally distributed estimator leads to underperformance.

Now consider the order-up-to level $S(\bar{d}_t, \sigma\tau, c_\alpha)$, where τ denotes $\sqrt{1+1/t}$. This order-up-to level is normally distributed with mean $\mu + c_\alpha\sigma\tau$ and variance ${}^{\sigma^2}\!/_t$, so it is a special case of Z with $w = \tau$ and $v = (w-1)c_\alpha$; note that w cannot be chosen freely, since the variance of the order-up-to level depends on the used estimator.

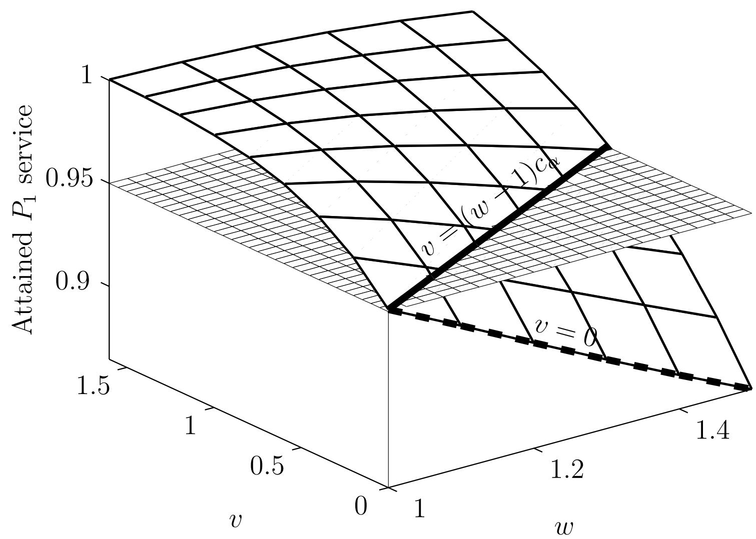


Figure 3.2: Attained P_1 service level ($\alpha = 0.95$) for estimator Z depending on v and w.

The performance of this order-up-to level ($S(\bar{d}_t, \sigma\tau, c_\alpha)$) is satisfactory. The sample mean $\bar{d}_t$ can be interpreted as a forecast of demand during the subsequent review period with forecast error variance $\sigma^2\tau^2$. So replacing the standard deviation of demand during review by the square root of the forecast error variance results in attaining the desired service. Note that σ is replaced by $\sigma\tau$, although this parameter is known. This may seem counterintuitive, but replacing σ by $\sigma\tau$ just increases the expected value of the order-up-to level; it does not change the variance of that level. Obviously, not knowing the expected demand can be compensated by putting a deliberate positive bias on the order-up-to level and that is exactly what happens when using $S(\bar{d}_t, \sigma\tau, c_\alpha)$.

3.3 P_2 service level criterion

This section focuses on the P_2 service criterion, also known as the fill rate. Using again the assumption that demand is normally distributed, the order-up-to level in this case should be (see, e.g., Silver et al., 1998)

$$S(\mu, \sigma, c_\beta) = \mu + c_\beta\sigma. \tag{3.4}$$

The safety factor c_β is given by

$$c_\beta = G^{-1}\left({}^{(1-\beta)}/_{\nu}\right), \tag{3.5}$$

where $G(\cdot)$ denotes the loss function of the standard normal distribution ($G(x) = \mathbb{E}\left[(Y-x)^+\right]$ with Y a standard normal random variable). Note that $G(x)$ is equal to $G(x) = \varphi(x) - x\Phi(-x)$, where $\varphi(\cdot)$ denotes the pdf of the standard normal distribution. As in the previous section, S without arguments denotes the correct order-up-to level. Furthermore, the service level attained by using an order-up-to level Z is

$$1 - \frac{\mathbb{E}\left[(D_R - Z)^+\right]}{\mathbb{E}\left[D_R\right]} = 1 - \frac{\mathbb{E}\left[(D_R - Z)^+\right]}{\mu}. \tag{3.6}$$

Note that if Z is constructed using estimates instead of the true values of the mean, standard deviation and coefficient of variation, it is a random variable. When using a P_2 service criterion, one needs values of μ and σ, and, in contrast to the P_1 criterion, also the coefficient of variation ν plays a role, as can be seen from (3.5). In practice, these parameters have to be estimated. The next two sections illustrate the effect of estimating μ and σ in the case where the correct safety factor is used, i.e., when ν is known. Section 3.3.3 considers the more realistic case where also the coefficient of variation needs to be estimated, and hence the safety factor as well.

3.3.1 Only expected demand unknown

This section assumes that only μ is unknown and thus that σ and ν are known. This is a purely theoretical assumption, since if σ and ν are known, μ can easily be determined. Yet, this case is interesting since even now the commonly used order-up-to level does not guarantee reaching the desired service level. If the sample mean is used to estimate μ, the order-up-to level is $S(\bar{d}_t, \sigma, c_\beta)$. Note that $S(\bar{d}_t, \sigma, c_\beta)$ is normally distributed with mean S and variance ${}^{\sigma^2}\!/t$. Using $S(\bar{d}_t, \sigma, c_\beta)$ does not result in attaining the desired service level β in the long run:

$$\begin{aligned} 1 - \frac{\mathbb{E}\left[\left(D_R - S(\bar{d}_t, \sigma, c_\beta)\right)^+\right]}{\mu} &= 1 - \frac{\sigma\tau G\left(\frac{c_\beta}{\tau}\right)}{\mu} < 1 - \nu\tau G\left(c_\beta\right) \\ &= 1 - \nu\tau\frac{1-\beta}{\nu} < \beta, \end{aligned} \tag{3.7}$$

where τ still denotes $\sqrt{1 + 1/t}$. The first inequality follows from the fact that the derivative of $G(x)$ is $-\Phi(-x) < 0$, so $G(x)$ is a strictly decreasing function. Note also that c_β has to be positive, or, $\beta > 1 - {}^{\nu}\!/\!\sqrt{2\pi}$. Again, this is not a limitation in practice, since the desired service level will be high. Only if $\nu < \sqrt{{}^{\pi}\!/50} \approx 0.2507$ the desired service level needs to be larger than 0.90 for the first inequality in (3.7) to hold.

Again, even in this simple case, using the unbiased estimator for S leads to a lower attained service level than β. We consider both the attained service levels ($\hat{\beta}$) and their relative deviations ($\delta_\beta(\hat{\beta})$). The relative deviation $\delta_\beta(\hat{\beta})$ is defined as

$$\delta_\beta(\hat{\beta}) = \frac{(1-\hat{\beta})-(1-\beta)}{(1-\beta)} = \frac{\beta-\hat{\beta}}{1-\beta},$$

where $\hat{\beta}$ is the attained service level and β the desired service level. This is a good performance measure, since it takes the height of the desired service level into account: if the service is 0.01 too low when the desired service level is 0.90, it will not be considered as bad as when the desired service level is 0.99. Note that when $\delta_\beta(\hat{\beta})$ is positive, $\hat{\beta}$ is less than β and we experience underperformance; when $\delta_\beta(\hat{\beta}) < 0$, there is overperformance. Overperformance might not be as bad as underperformance, but it implies that too many items are in inventory and we could reach the desired service level with less inventory, hence with less costs. Five different attained service levels and their relative deviations are discussed in the remainder of this chapter; these attained service levels and there notation is listed in Table 3.1.

$\hat{\beta}_0$:	Attained service when μ unknown, but σ and ν known (Section 3.3.1);
$\hat{\beta}_1$:	Attained service when μ and σ unknown, but ν known (Section 3.3.2);
$\hat{\beta}_2$:	Attained service when μ, σ, and ν unknown (Section 3.3.3);
$\hat{\beta}_3$:	Attained service when using correction function $\hat{\kappa}_\sigma(\nu, t, \beta)$ (Section 3.3.4);
$\hat{\beta}_4$:	Attained service when using correction function $\hat{\kappa}_s(\nu, t, \beta)$ (Section 3.3.4).

Table 3.1: The attained service levels discussed in Section 3.3 and their notation.

The attained service and its relative deviation assuming that only μ is unknown are shown in Figure 3.3; note that the x- and y-axes in the right graphs are reversed compared to the left graphs. This figure shows that the underperformance is larger when t is smaller (ceteris paribus, c.p. for short), when ν is larger (c.p.) and when β is larger (c.p.; underperformance measured through $\delta_\beta(\hat{\beta}_0)$). When t is smaller, the variance of the order-up-to level is larger, hence, intuitively the underperformance should be larger too. The same line of reasoning applies to ν being larger; also in that case the variance of the order-up-to level is larger and thus the underperformance is larger. If

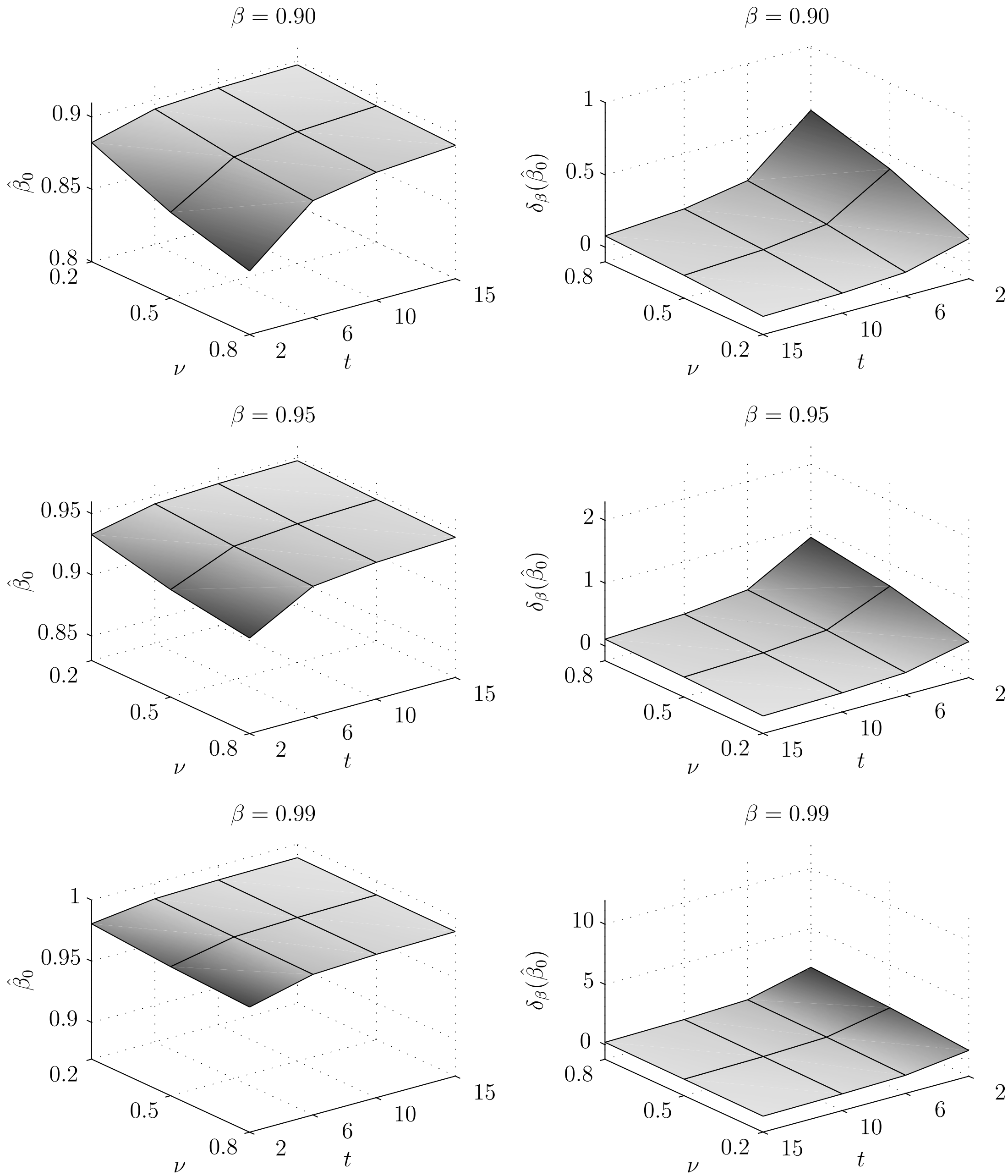


Figure 3.3: Attained service ($\hat{\beta}_0$) and relative deviation ($\delta_\beta(\hat{\beta}_0)$) if only μ unknown.

β is larger, $1 - \beta$ is smaller and the same absolute deviation is worse if β is larger, hence the underperformance, measured through the relative deviation, is larger. The extreme attained service levels for all three desired service levels are displayed in Table 3.2. One can clearly see that the deviations from the desired service level cannot be neglected, since they can become as large as 0.0561 ($0.90 - 0.8439$), which

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_0)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_0)$)	Mean attained service ($\delta_\beta(\hat{\beta}_0)$)
$\beta = 0.90$	0.8439 (0.561)	0.8974 (0.026)	0.8838 (0.162)
$\beta = 0.95$	0.9080 (0.840)	0.9475 (0.050)	0.9372 (0.256)
$\beta = 0.99$	0.9721 (1.790)	0.9888 (0.120)	0.9844 (0.560)

Table 3.2: Extreme deviations for $\beta \in \{0.90, 0.95, 0.99\}$ when only μ unknown.

is more than 5 percent points below the desired service.

In case of the P_1 criterion, the problem of underperformance was solved by replacing the standard deviation σ by the square root of the forecast error variance, $\sigma\tau$. So we might apply the same adjustment in this case. Note however that σ has to be replaced twice: once explicitly and once implicitly, as the safety factor c_β depends on σ via ν. Now let us denote the new safety factor by c_β^τ:

$$c_\beta^\tau = G^{-1}\left(\frac{1-\beta}{\nu\tau}\right). \tag{3.8}$$

Obviously $c_\beta^\tau > c_\beta$. Upwards biasing the safety factor has been mentioned in the literature; see Section 3.1. However, to our knowledge, applying the factor τ both to the explicit and the implicit standard deviation has never been mentioned. Now consider the attained service level when using the order-up-to level $S(\bar{d}_t, \sigma\tau, c_\beta^\tau)$:

$$1 - \frac{\mathbb{E}\left[\left(D_R - S(\bar{d}_t, \sigma\tau, c_\beta^\tau)\right)^+\right]}{\mu} = 1 - \nu\tau G\left(c_\beta^\tau\right) = 1 - \nu\tau\frac{1-\beta}{\nu\tau} = \beta. \tag{3.9}$$

So using $S(\bar{d}_t, \sigma\tau, c_\beta^\tau)$ indeed results in attaining the desired service level in the long run.

3.3.2 Expected demand and its variance unknown (ν known)

Now assume that only ν is known, so σ has to be estimated too. We choose to use the sample standard deviation s_t. One could also use the known ν and the estimate of μ to estimate the value of σ, but since the next step is to assume that ν is also unknown, we do not follow that approach. The sample standard deviation is defined as $s_t = \sqrt{\frac{1}{t-1}\sum_{i=1}^t \left(d_i - \bar{d}_t\right)^2}$. Substituting this estimator yields the order-up-to level $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$. Consider the expected value of this order-up-to level and note that s_t^2

is an unbiased estimator of σ^2, so $\mathbb{E}\left[s_t^2\right] = \sigma^2$. However, from Jensen's Inequality (see, e.g., Mittelhammer, 1996) it follows that $\mathbb{E}\left[s_t\right] < \sigma$. This implies

$$\begin{aligned}\mathbb{E}\left[S(\bar{d}_t, s_t\tau, c_\beta^\tau)\right] &= \mathbb{E}\left[\bar{d}_t + s_t c_\beta^\tau \tau\right] \\ &= \mathbb{E}\left[\bar{d}_t\right] + c_\beta^\tau \tau \mathbb{E}\left[s_t\right] < \mu + \sigma c_\beta^\tau \tau = \mathbb{E}\left[S(\bar{d}_t, \sigma\tau, c_\beta^\tau)\right], \end{aligned} \tag{3.10}$$

if $\beta > 1 - {}^{\tau\nu}/\sqrt{2\pi}$. This constraint is even less restricting compared to (3.7), since ${}^{\tau\nu}/\sqrt{2\pi} > {}^{\nu}/\sqrt{2\pi}$. Thus the expected value of $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$ is lower than the expected value of the order-up-to level in the case where σ is known. Furthermore, σ being unknown results in extra variability, so the performance of $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$ will probably be lower than desired.

Since we want to quantify the underperformance, we perform simulation runs for different values of t ($t \in \{2, 6, 10, 15\}$), ν ($\nu \in \{0.2, 0.5, 0.8\}$) and β ($\beta \in \{0.90, 0.95, 0.99\}$). For each combination of t and ν we randomly generated $n = 1,000,000$ samples of $t+1$ normally distributed observations with mean $1/\nu$ and standard deviation 1. The mean μ and standard deviation σ need not to be varied, since the performance does not depend on these parameters separately (see Appendix C.1). We estimate μ and σ using $\bar{d}_t$ and s_t and the order-up-to levels are determined using $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$. Subsequently the attained service, denoted by $\hat{\beta}_1$ (see Appendix C.1 for the definition), is estimated. The results are shown in Figure 3.4. Simulation shows that indeed the desired service level is not attained, except for the cases that $\beta = 0.90$ and $\nu = 0.2$. That is no surprise, since the constraint we needed for (3.10) to hold is violated in these cases. We furthermore find that the underperformance ($\beta - \hat{\beta}_1$) is larger if t is smaller and ν is larger. This is exactly what we expected to happen, since if ν is larger the variability of the order-up-to level is larger. Also if t is smaller the variability of the order-up-to level is larger. Hence, the expected amount of backorders is larger and thus the achieved service is less. Finally, the relative underperformance ($\delta_\beta(\hat{\beta}_1)$) is larger if β is larger. The most extreme deviations are listed in Table 3.3.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_1)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_1)$)	Mean attained service ($\delta_\beta(\hat{\beta}_1)$)
$\beta = 0.90$	0.8252 (0.748)	0.9008 (-0.008)	0.8875 (0.125)
$\beta = 0.95$	0.8625 (1.750)	0.9494 (0.012)	0.9328 (0.344)
$\beta = 0.99$	0.9045 (8.550)	0.9884 (0.160)	0.9719 (1.810)

Table 3.3: Extreme deviations for $\beta \in \{0.90, 0.95, 0.99\}$ when μ and σ unknown.

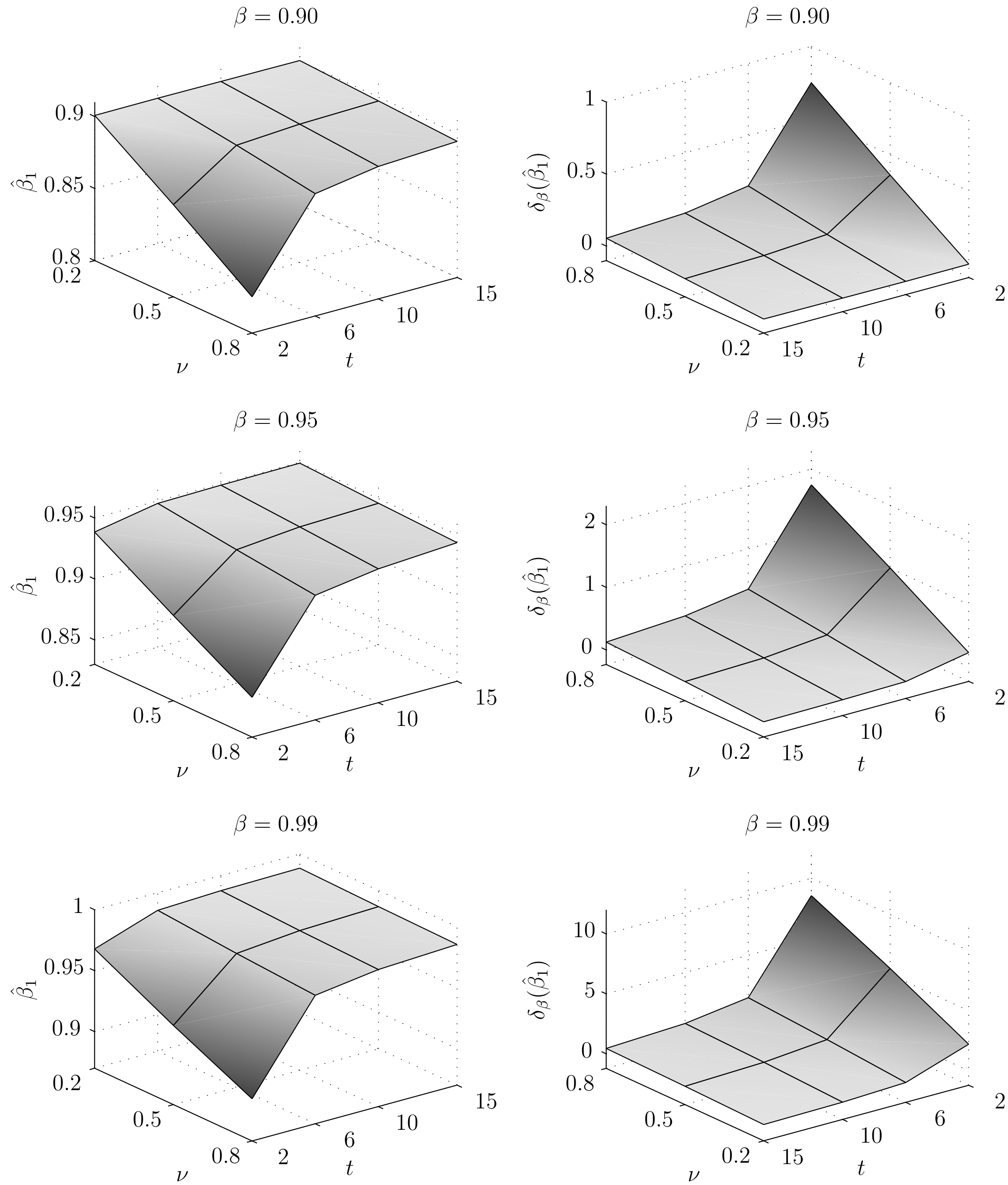


Figure 3.4: Attained service ($\hat{\beta}_1$) and relative deviation ($\delta_\beta(\hat{\beta}_1)$) if μ and σ unknown.

Since it is difficult to consider this case analytically and it is not of practical interest (as ν is assumed to be known), we continue with the most practical case, in which all parameters are unknown.

3.3.3 Expected demand and its variance unknown (ν unknown)

In this section the demand is normally distributed with unknown mean and standard deviation, and the coefficient of variation is also unknown. Estimates are used for all these unknown parameters, namely $\bar{d}_t$, s_t and $\hat{\nu}_t = s_t/\bar{d}_t$. Since the safety factor c^τ_β depends on ν, this factor has also to be estimated; $\hat{c}^\tau_\beta$ denotes this estimate and it is defined as

$$\hat{c}^\tau_\beta = \begin{cases} G^{-1}\left(\frac{1-\beta}{\hat{\nu}_t\tau}\right) & \text{if } \hat{\nu}_t > 0 \\ -\frac{1}{\hat{\nu}_t\tau} & \text{otherwise.} \end{cases} \tag{3.11}$$

Note that ν is simply replaced by $\hat{\nu}_t$, if it is possible to do so. Since $\bar{d}_t$ may be negative (the demand values are generated using a normal distribution), $\hat{\nu}_t$ can be negative as well. In that case the function $G^{-1}(\cdot)$ has no outcome, as its domain is strictly positive. If $\bar{d}_t$ is negative, it means that the demand in the next period is forecasted to be negative. So inventory is not needed in that case and hence $\hat{c}^\tau_\beta$ is chosen in such a way that the resulting order-up-to level equals zero.

For this case the order-up-to level is even more complicated than in the previous section, so again it is not possible to get analytical results. Therefore simulation is applied; first to estimate the attained service level when using order-up-to level $S(\bar{d}_t, s_t\tau, \hat{c}^\tau_\beta)$ and second to find a correction to that order-up-to level that would assure that the desired service is reached more closely. The P_2 service using $S(\bar{d}_t, s_t\tau, \hat{c}^\tau_\beta)$ is estimated with help of $n = 1,000,000$ simulation runs for each combination of t, β and ν. Note that again the attained service depends only on ν and not on μ and σ separately; see Appendix C.1 for further details. This simulation is performed in the same way as described in the previous section; the only difference is that $\hat{c}^\tau_\beta$ is used instead of c^τ_β in determining the order-up-to levels. The results are based on the same samples as used in the previous section; the attained service level ($\hat{\beta}_2$) and relative

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_2)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_2)$)	Mean attained service ($\delta_\beta(\hat{\beta}_2)$)
$\beta = 0.90$	0.8005 (0.995)	0.8998 (0.002)	0.8838 (0.162)
$\beta = 0.95$	0.8345 (2.310)	0.9486 (0.028)	0.9280 (0.440)
$\beta = 0.99$	0.8733 (11.670)	0.9877 (0.230)	0.9673 (2.270)

Table 3.4: Extreme deviations for $\beta \in \{0.90, 0.95, 0.99\}$ when μ, σ and ν unknown.

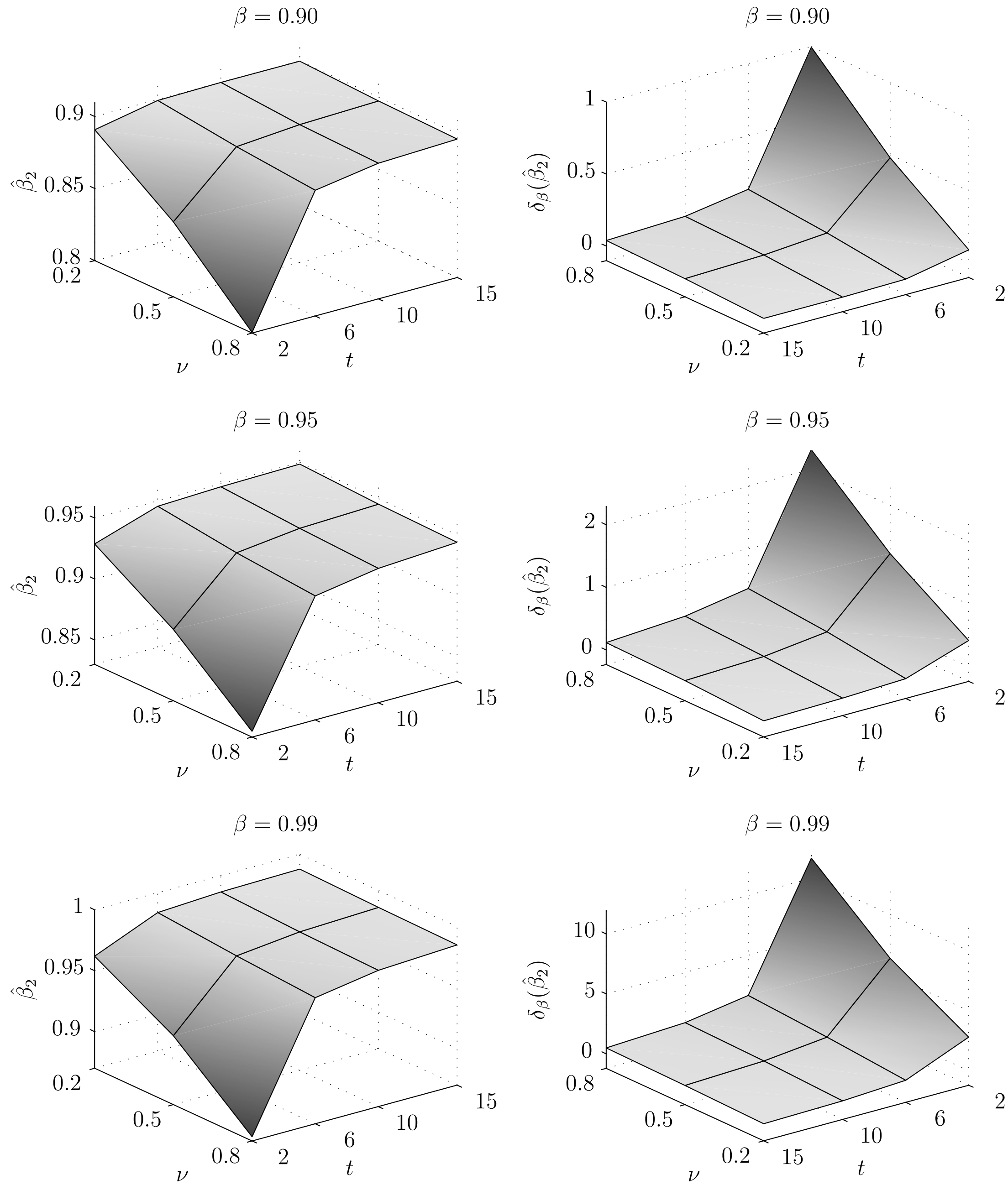


Figure 3.5: Attained service ($\hat{\beta}_2$) and relative deviation ($\delta_\beta(\hat{\beta}_2)$) if μ, σ and ν unknown.

deviation ($\delta_\beta(\hat{\beta}_2)$) are shown graphically in Figure 3.5; Table 3.4 lists the extreme deviations. In most cases the performance is indeed worse, as expected. In a few cases ($t = 10, 15$, $\nu = 0.8$, and $\beta = 0.90, 0.95$), the performance of the $S(\bar{d}_t, s_t\tau, \hat{c}_\beta^\tau)$ is slightly better than the performance of $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$. Furthermore, the same overall results as in Section 3.3.2 appear: the underperformance increases as t decreases and as ν increases. The relative underperformance also increases as β increases.

3.3.4 Correction of the order-up-to level

Now the underperformance is quantified in Section 3.3.3, we also want to find a correction for $S(\bar{d}_t, s_t\tau, \hat{c}^{\tau}_{\beta})$, such that the desired service is reached or at least approached more closely. This correction depends on ν, t and β and is denoted by $\kappa\,(\nu, t, \beta)$. Such a correction function can be useful in practice, since it provides inventory managers with a simple tool to improve their easy-to-understand order-up-to levels. After considering several options to correct the order-up-to level, a dimensionless correction factor is determined in order to provide an upwards bias to $S(\bar{d}_t, s_t\tau, \hat{c}^{\tau}_{\beta})$. The corrected order-up-to level is $S(\bar{d}_t + \kappa\,(\nu, t, \beta)\, s_t, s_t\tau, \hat{c}^{\tau}_{\beta})$. Note that this is not applicable in practice, since ν is unknown. However, replacing ν by $\hat{\nu}_t$ does improve the attained service, as is shown in Figures 3.8 and 3.9.

Simulation is used to estimate the correction needed for various values of ν, t and β. The two simulations performed earlier used only a limited number of values for the three parameters; more values are needed to be able to estimate the correction using ν, t, and β. So a new simulation is performed in which $n = 100,000$ samples of $t+1$ observations are generated for each combination of t, ν, and β; (3.12) provides the values of the three parameters used in the simulation:

$$\begin{aligned}
\mathcal{T} &= \{2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20\} \\
\mathcal{V} &= \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\} \\
\mathcal{B} &= \{0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99\}.
\end{aligned} \tag{3.12}$$

The range of the values do not differ much from the range used in the earlier simulations (t between 2 and 15; ν between 0.2 and 0.8; β between 0.90 and 0.99). The size of the underperformance will not become much larger; only when $\nu = 0.9$ and $\nu = 1$ and t small the differences are larger than the values reported in Tables 3.2–3.4.

The number of generated samples is smaller, since we need considerably more simulations, which are time-consuming; the complete simulation needed more than one week to finish. For each simulation run j ($j = 1, \ldots, n$) the estimated order-up-to level, denoted by S_j, is calculated using the first t observations and the $(t+1)$th observation d_j is used to quantify the backlogged demand that occurred. The correction needed can be determined using either the true value σ or the sample standard deviations s_{tj} ($j = 1, \ldots, n$) found in the simulation. On the one hand using the true value might be better, since it is the *true* value. On the other hand, a correction is needed since an estimate is used instead of the true value. It is difficult to decide a priori which would result in a better correction formula, so both corrections, denoted by k_σ and k_s respectively, are determined, by solving the following two equations:

$$\sum_{j=1}^{n} (d_j - (S_j + k_\sigma \sigma))^+ = (1-\beta) \sum_{j=1}^{n} d_j, \tag{3.13}$$

$$\sum_{j=1}^{n} (d_j - (S_j + k_s s_{tj}))^+ = (1-\beta) \sum_{j=1}^{n} d_j. \tag{3.14}$$

The values for k_σ and k_s for all combinations of ν, t and β can be found by solving this equation numerically using binary search. Note that the left hand sides of (3.13) and (3.14) are decreasing in k_σ and k_s, so there is one value for k_σ and one for k_s for which equality holds. The corrections are denoted by $k_i(\nu, t, \beta)$ ($i \in \{\sigma, s\}$). The corrections for different values of ν, t and β are depicted in Figures 3.6 and 3.7.

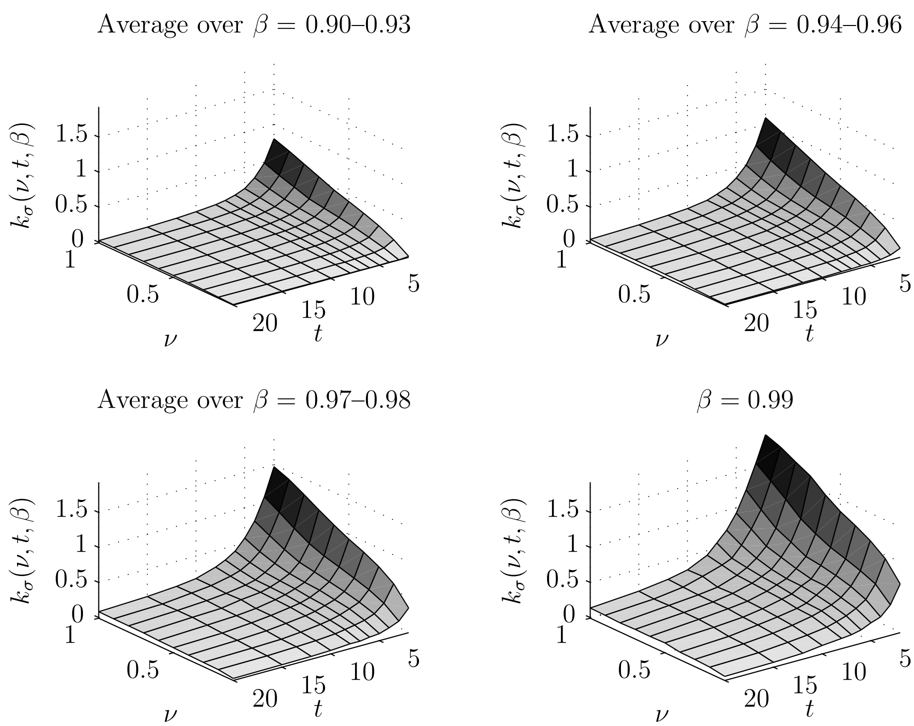


Figure 3.6: Corrections $k_\sigma(\nu, t, \beta)$ of the order-up-to level $S(\bar{d}_t, s_t\tau, \hat{c}^\tau_\beta)$.

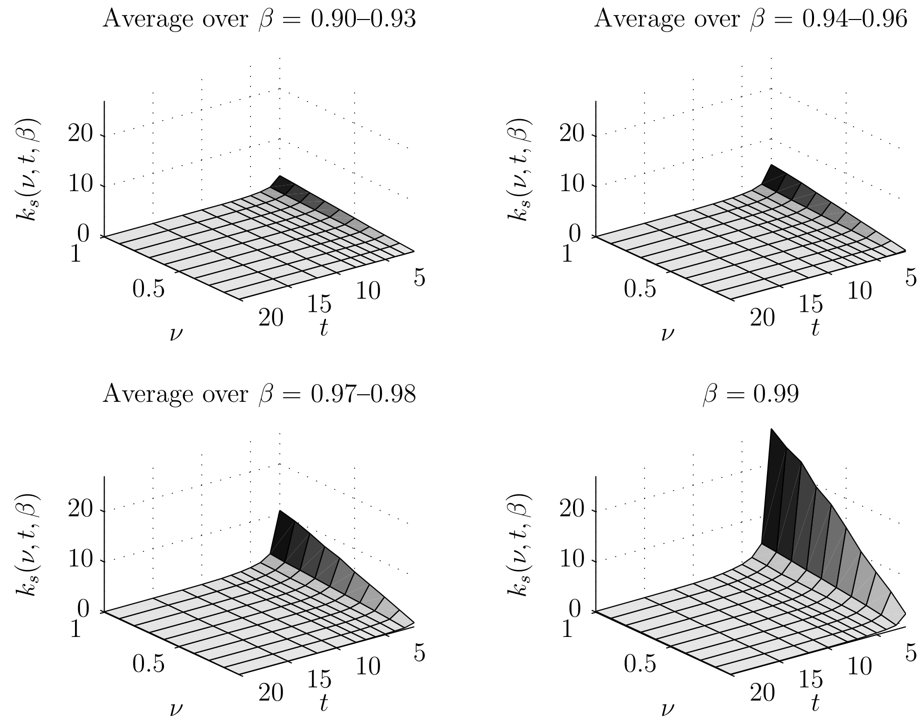


Figure 3.7: Corrections $k_s\,(\nu, t, \beta)$ of the order-up-to level $S(\bar{d}_t, s_t\tau, \hat{c}^{\tau}_{\beta})$.

For the first three graphs in Figures 3.6 and 3.7 the corrections needed for different values of β are averaged, since they do not differ much. One sees that if ν increases, the correction needs to be larger. The same is true if t decreases and if β increases. These conclusions correspond to the results in Section 3.3.3.

The simulation resulted in some values $k_\sigma\,(\nu, t, \beta)$ and $k_s\,(\nu, t, \beta)$, but a function that is able to estimate these values is more practical. Therefore a regression technique we call nested regression is applied, following the method described by Strijbosch and Moors (1999). This method is described elaborately in Appendix B.

The estimation process is split into three steps:

1. Choose one of the three parameters (denoted by q_1), and use linear regression to estimate the correction needed depending on that parameter for each combination of the other two parameters;

2. Choose one of the remaining two parameters (q_2) and regress that parameter on the coefficients found in Step 1 for each value of the third parameter;

3. Regress the third parameter (q_3) on the coefficients found in Step 2.

Finally the regression equations are combined in order to get a correction function.

A different ordering of choosing the parameters could influence the result, hence all six orderings are examined, both for $k_\sigma(\nu, t, \beta)$ and $k_s(\nu, t, \beta)$. Example B.2 in Appendix B shows how $k_s(\nu, t, \beta)$ is constructed in detail. The fit, determined by R^2 (coefficient of determination), of all possible orderings of adding the parameters and for both $k_\sigma(\nu, t, \beta)$ and $k_s(\nu, t, \beta)$ are given in Table 3.5. So using $k_s(\nu, t, \beta)$ instead

q_1	q_2	q_3	R^2 of $k_\sigma(\nu, t, \beta)$	R^2 of $k_s(\nu, t, \beta)$
t	ν	β	**0.9897**	0.9960
t	β	ν	0.9894	0.9962
ν	t	β	0.9828	**0.9963**
ν	β	t	0.9872	0.9889
β	t	ν	0.9876	0.9958
β	ν	t	0.9840	0.9900

Table 3.5: R^2 found for different orderings of choosing parameters.

of $k_\sigma(\nu, t, \beta)$ results in slightly better values of R^2, although the R^2s for both corrections are very good. Next a simulation is performed to determine which correction provides the best performance. Again the samples of size $n = 1,000,000$ are used (see Section 3.3.2), now to determine the performance of $S(\bar{d}_t + \hat{\kappa}_\sigma(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$ and $S(\bar{d}_t + \hat{\kappa}_s(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$. The function $\hat{\kappa}_\sigma(\hat{\nu}_t, t, \beta)$ is based on $k_\sigma(\nu, t, \beta)$ determined using $q_1 = t$, $q_2 = \nu$ and $q_3 = \beta$, while $\hat{\kappa}_s(\hat{\nu}_t, t, \beta)$ is based on $k_s(\nu, t, \beta)$ determined using $q_1 = \nu$, $q_2 = t$ and $q_3 = \beta$; the corresponding R^2s are printed in boldface in Table 3.5. Also the performance of these order-up-to levels is independent of σ and μ; see Appendix C.1.

Figure 3.8 shows the attained service levels ($\hat{\beta}_3$) and the corresponding relative deviations ($\delta_\beta(\hat{\beta}_3)$) when using $\hat{\kappa}_\sigma(\cdot)$. Table 3.6 lists the extreme deviations using this correction function.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_1)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_3)$)	Mean attained service ($\delta_\beta(\hat{\beta}_3)$)
$\beta = 0.90$	0.8449 (0.551)	0.9005 (-0.005)	0.8917 (0.083)
$\beta = 0.95$	0.8882 (1.236)	0.9497 (0.006)	0.9400 (0.200)
$\beta = 0.99$	0.9332 (5.680)	0.9900 (0.000)	0.9796 (1.040)

Table 3.6: Extreme deviations for $\beta \in \{0.90, 0.95, 0.99\}$ using correction $\hat{\kappa}_\sigma(\hat{\nu}_t, t, \beta)$.

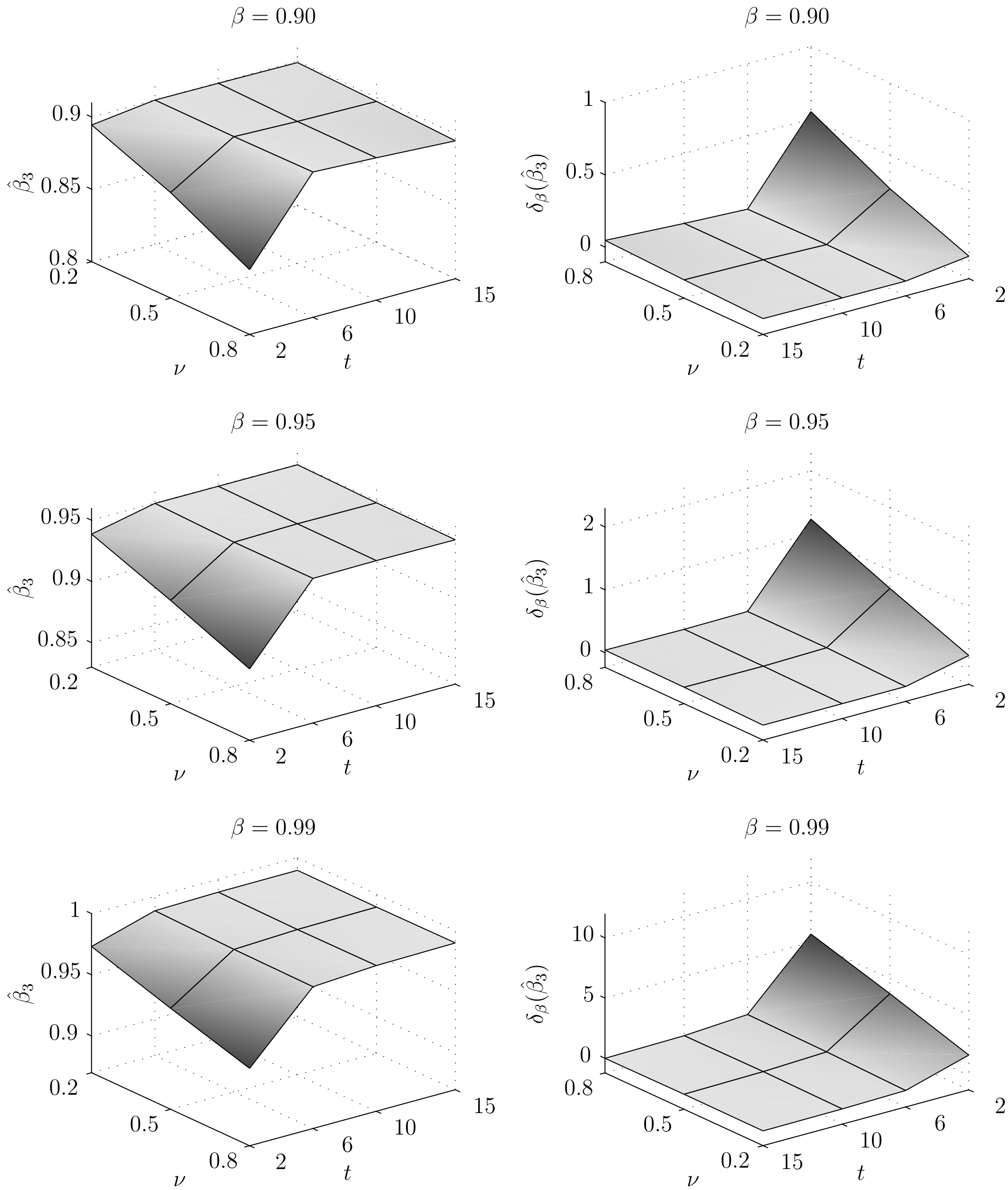


Figure 3.8: Attained service ($\hat{\beta}_3$) and relative deviation ($\delta_\beta(\hat{\beta}_3)$) using correction $\hat{\kappa}_\sigma\,(\hat{\nu}_t, t, \beta)$.

The resulting attained service levels ($\hat{\beta}_4$) when using $\hat{\kappa}_s\,(\cdot)$ are displayed in Figure 3.9, together with the relative deviations ($\delta_\beta(\hat{\beta}_4)$). The extreme deviations when using the second correction function are shown in Table 3.7. Comparing the results

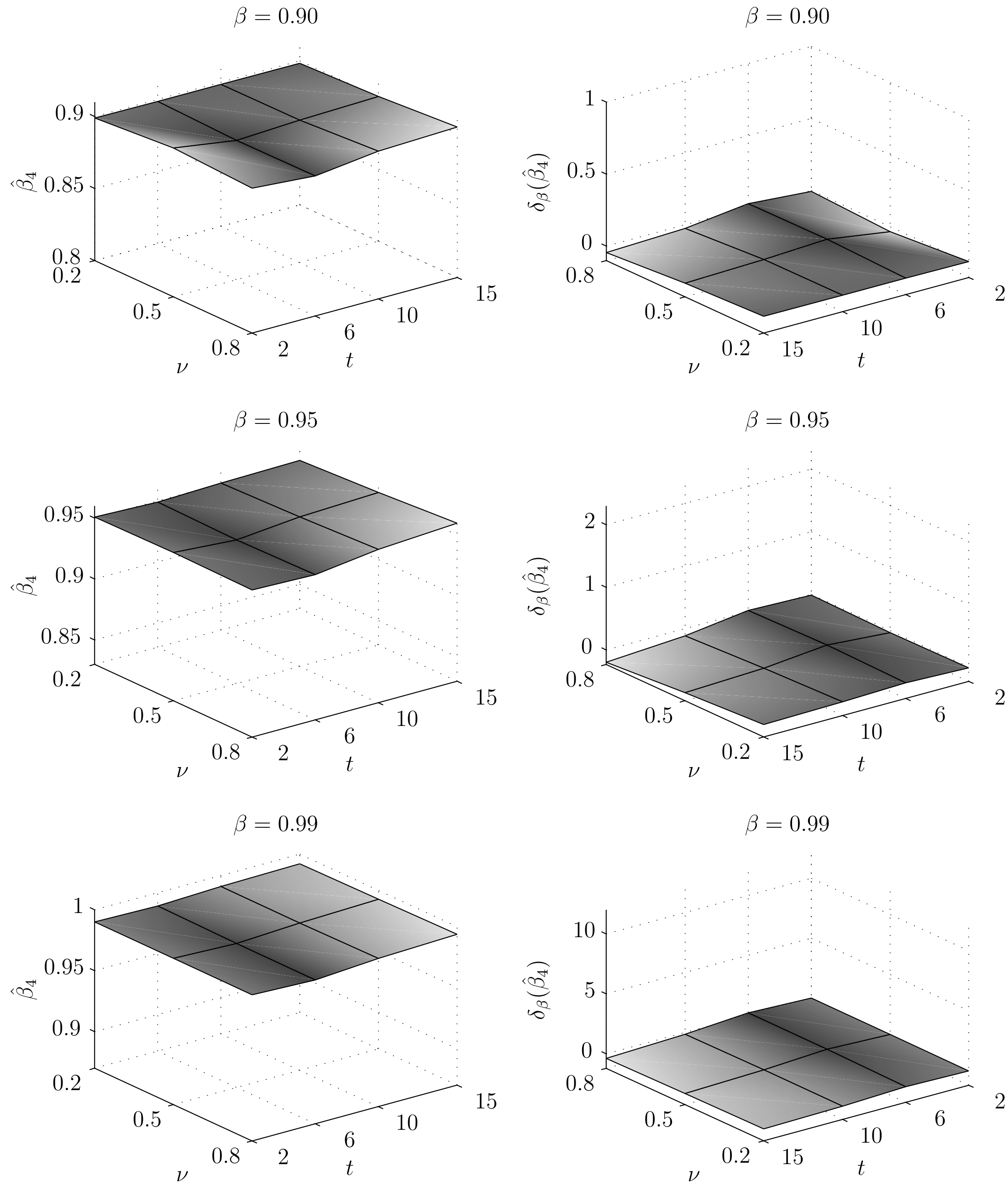


Figure 3.9: Attained service ($\hat{\beta}_4$) and relative deviation ($\delta_\beta(\hat{\beta}_4)$) using correction $\hat{\kappa}_s(\hat{\nu}_t, t, \beta)$.

of using a correction, either $\hat{\kappa}_\sigma(\cdot)$ or $\hat{\kappa}_s(\cdot)$, to not using a correction, we see that the achieved service level is closer to the desired one in most cases (33 out of 36 for $\hat{\kappa}_\sigma(\cdot)$ and 28 out of 36 for $\hat{\kappa}_s(\cdot)$). So we may conclude that using a correction improves the attained service level.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_4)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_4)$)	Mean attained service ($\delta_\beta(\hat{\beta}_4)$)
$\beta = 0.90$	0.8968 (0.032)	0.9045 (-0.045)	0.8999 (0.001)
$\beta = 0.95$	0.9479 (0.042)	0.9597 (-0.194)	0.9520 (-0.040)
$\beta = 0.99$	0.9879 (0.210)	0.9937 (-0.370)	0.9906 (-0.060)

Table 3.7: Extreme deviations for $\beta \in \{0.90, 0.95, 0.99\}$ using correction $\hat{\kappa}_s(\hat{\nu}_t, t, \beta)$.

Next we compare the corrections. If $\hat{\kappa}_\sigma(\cdot)$ is used, the attained service level is closer to the desired one in 19 out of 36 cases compared to $\hat{\kappa}_s(\cdot)$. Based on this result one may think that $\hat{\kappa}_\sigma(\cdot)$ is the better correction. However, if t is small, $S(\bar{d}_t + \hat{\kappa}_\sigma(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$ performs far worse compared to $S(\bar{d}_t + \hat{\kappa}_s(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$. The first order-up-to level does not reach the desired service if t is small, while it does approximately if t is large. The performance of $S(\bar{d}_t + \hat{\kappa}_s(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$ hardly depends on the values of t, ν and β, so $\hat{\kappa}_s(\hat{\nu}_t, t, \beta)$ seems to be the best correction. This can be explained as follows: $\hat{\kappa}_\sigma(\cdot)$ is based on the true value of σ, which does not depend on t, while $\hat{\kappa}_s(\cdot)$ is based on the estimated value of σ (s_t). This estimate and its variation do depend on t, so the effect of t is taken into account in a more elaborate way in case of $\hat{\kappa}_s(\cdot)$. Finally, the formula for $\hat{\kappa}_s(\nu, t, \beta)$ is:

$$\begin{aligned} &\hat{\kappa}_s(\nu, t, \beta) = \\ &-0.0669 + 0.00305 \cdot (1-\beta)^{-0.95} + \left(-185.124 - 6.359 \cdot (1-\beta)^{-1.00}\right) t^{-9.17} \\ &\quad + \left[0.335 - 5.671 \cdot (1-\beta)^{1.41} + \left(-3.841 + 4.541 \cdot (1-\beta)^{-1.03}\right) t^{-4.19}\right] \nu^{0.90}. \end{aligned} \tag{3.15}$$

3.4 Summary results

This chapter has investigated a common approach in inventory management for dealing with the unknown distribution of demand. This approach is to assume a distribution, estimate its parameters using historical demand information, and replace the parameters in the theoretically correct inventory model by its estimates.

We have assumed that the demand during review is truly normally distributed. In steps the information about the mean and variance is reduced. First, only the mean is assumed to be unknown and using the common approach results in the order-up-to levels $S(\bar{d}_t, \sigma, c_\alpha)$ for a P_1 criterion and $S(\bar{d}_t, \sigma, c_\beta)$ for a P_2 criterion. Both order-up-to levels do not ensure that the desired service is reached. This can be resolved by replacing the standard deviation σ with the square root of the forecast error variance $\sigma\tau$, where $\tau = \sqrt{1 + 1/t}$.

Second, also the standard deviation becomes unknown, although the coefficient of variation is assumed to be known. This case is less tractable and, therefore, only the more interesting P_2 criterion is considered for this case. The order-up-to level using the correction factor τ becomes $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$, for which we have shown that the expected value is too low. Furthermore, simulation has shown that the desired service is indeed not reached when using this order-up-to level. Underperformance ranges from 0.0875 to 0.0006 (in one case a small overperformance of 0.0008 is obtained); the relative deviation ranges from 8.550 to -0.008).

Finally, also the coefficient of variation ν is unknown and in that case only simulation is used to find that the performance is again worse compared to the case when ν is known. Underperformance ranges from 0.1167 to 0.0002; the relative deviation ranges from 11.670 to 0.002.

We have developed a correction for the order-up-to level with the help of simulation and linear regression. Using this correction, which is a function of ν, β and t, ensures that the desired service is reached closely in general: the largest deviations are 0.0032 below the desired service and 0.0097 above; the relative deviation ranges from 0.042 to -0.370.

It can be concluded that simply replacing the parameters in a theoretical correct inventory control model by its estimates results in underperformance, even if the true distribution belongs to the assumed distribution family. With a simple correction the achieved service can be improved and using the correction function $\hat{\kappa}_s(\nu, t, \beta)$ results in closely reaching the desired service. This correction function provides inventory managers with a relatively simple tool that improves the service levels of their SKUs and can be implemented without much effort.

Chapter 4

Gamma demand with unknown demand parameters

Inventory models need information about the demand distribution. In practice, this information is not known with certainty and has to be estimated with often relatively few historical demand observations. Using these estimates leads to underperformance. This chapter focuses on gamma distributed demand and an (R, S) inventory control policy, where the order-up-to level satisfies a service equation. Under this policy the underperformance is quantified analytically under strong assumptions and with help of simulation if these assumptions are relaxed. The attained service level can be improved with help of the analytical results. Using simulation and nested regression further improvements can be obtained. These two improvements lead to (almost) reaching the desired service level. Finally, the method developed in this chapter is applied to real demand data using simulation. This chapter is based on Janssen et al. (2007).

4.1 Introduction

Inventory control models need information about the demand distribution. These models are developed assuming that all the information they need (e.g., moments, family of distribution, parameters) is known with certainty. However, in practice often relatively few historical demand observations are known and these are used to estimate the demand distribution characterization that is needed.

In this chapter the family of distributions to which the demand belongs, is assumed to be known, but its parameters are not. Hence, estimates are needed to use the inventory model and the effect of using these estimates is studied. In Chapter 3 the effects of using estimates with normally distributed demand is studied. Although the normal distribution is commonly assumed in inventory control, it certainly has some problems. According to Burgin (1975) demand distributions generally only exist for nonnegative values of demand and the shape of the density function changes from monotonic decreasing (low mean demand) via a unimodal distribution that is skewed to the right to a normal type distribution that is truncated at zero (high mean demand). A normal distribution does not fit all these criteria: the probability that a normally distributed variable can be negative, is nonnegligible (more than 1%) if the coefficient of variation is larger than 0.43; further, the normal distribution is symmetric. The gamma distribution does fit the criteria of Burgin (1975), since it is nonnegative and the value of the shape parameter can be adjusted to get all three forms described. The gamma distribution also has some nice properties which makes it relatively easy to work with, although maybe not as easy as the normal distribution, and that is probably why the normal distribution is used so often in literature and in practice.

The gamma distribution has proven its worth. Watson (1987) considers an Erlang distribution (i.e., a special case of the gamma distribution) and studies the effect forecasting has on attaining the desired service level using simulation. Note that the Erlang assumption implies that demand during lead time has a relatively small coefficient of variation, hence demand cannot be highly variable, which limits the applicability of Watson (1987). Furthermore, he does restrict his research to intermittent demand. Segerstedt (1994) develops another inventory control policy, which also uses gamma distributions. Yeh (1997) slightly adapts this policy to implement it in an electronics industrial company. Both mention that parameters in the model need to be estimated, but they do not show the effects of doing this. The consultancy firm Involvation has applied the gamma distribution in their stock control software and its customers are satisfied with the achieved improvements. This consultancy firm generously provided demand data of one of their customers, the Dutch Ministry of Defence, which is used to test the method developed in this chapter.

The demand is assumed to follow a gamma process, i.e., the demand during a period of length ℓ, denoted by D_ℓ, has a gamma distribution with shape parameter $\ell\rho$ and scale parameter θ, or $D_\ell \sim \Gamma(\ell\rho, \theta)$ for short. If $\ell = 1$ demand is denoted by D. Also, demands during disjoint time intervals are independent.

Demand is assumed to be stationary for $t + L + 1$ periods, which means that the actual demand during the first t periods can be used to estimate the demand during the last $L+1$ periods. This method of forecasting is known as moving average and it is used because either one has only few observations or one wants to account for nonstationarity. In the latter case one cannot use all the demand observations in a large set of historical demand data, since the demand pattern has changed over time and earlier demand observations might not reflect the current demand pattern. The moving average tries to take nonstationarity into account by only considering the last t demand observations, for which we assume that they do reflect the demand as it will occur in the coming $L+1$ periods. Besides that, moving average also allows us to find some analytical results.

Furthermore, an (R, S) inventory control policy is used with $R = 1$ (without loss of generality). This policy states that every R periods the inventory position is reviewed and replenished up to S. The order is then delivered after a fixed and deterministic lead time L. It is possible that an order is negative, since the order-up-to level is determined using estimates: if some periods with high demand are followed by some periods with low demand, the order-up-to level is high first, but low later and then a negative order can arise. We assume that products can be sent back to the supplier in that case. If products cannot be sent back, the attained service levels will be higher.

The order-up-to level S is chosen such that the required service level is reached. This chapter considers both the P_1 and P_2 service level. The demand that cannot be satisfied immediately, is backlogged.

The remainder of the chapter consists of three parts. Section 4.2 considers the P_1 service criterion. The theoretically correct order-up-to level is provided and the impact of using estimates for the unknown parameters is investigated. We prove that underperformance always occurs under the assumption that demand is exponentially distributed and that it occurs when the desired service level is over 50% in case of Erlang demand with a known shape parameter. Relaxing these assumptions leads to intractable results and at this point simulation is used to show that underperformance occurs for values of the desired service level that are commonly used in practice. With help of simulation a correction to the order-up-to level is found such that the desired service level is reached and regression techniques are applied to estimate the relation between the correction needed and the parameters of the model. Using the regression equation to correct the order-up-to level results in reaching the desired service level more closely. Section 4.3 has approximately the same structure, but focuses on the

P_2 service criterion. It provides the theoretically correct order-up-to level under the assumption that the complete demand distribution is known. Further, the effects of using estimates are discussed. We prove that the order-up-to levels under the P_1 and the P_2 criterion are equal in case of exponentially distributed demand and hence the results of the P_1 criterion can be adopted. For the remainder of this section simulation is used to show that underperformance occurs if the desired service level is relatively high. Further, a correction to the estimated order-up-to level is provided; using this correction reduces the underperformance significantly. Section 4.4 uses the order-up-to levels developed in Sections 4.2 and 4.3 on actual demand data; large improvements are obtained. Section 4.5 concludes this chapter with a short summary of the results.

4.2 P_1 service level criterion

This section focuses on the P_1 service level criterion, so we need an order-up-to level such that $\mathbb{P}\left(D_{1+L} \leq S\right) = \alpha$. Furthermore, we know that $D_{1+L} \sim \Gamma((1+L)\rho, \theta)$. Define $\widetilde{\rho} := (1+L)\rho$. If the parameters of a gamma distribution are known, the order-up-to level is easily determined:

$$\alpha = F_{\widetilde{\rho},\theta}(S) \qquad \Leftrightarrow \qquad S = F^{-1}_{\widetilde{\rho},\theta}(\alpha) \qquad \Leftrightarrow \qquad S = \theta F^{-1}_{\widetilde{\rho},1}(\alpha). \tag{4.1}$$

The function $F_{\rho,\theta}$ ($F^{-1}_{\rho,\theta}$) is the distribution function (inverse distribution function) of a gamma distribution with parameters ρ and θ. In the next section also $f_{\rho,\theta}$ is used, which denotes the corresponding density function.

4.2.1 Using estimates in determining the order-up-to level

The order-up-to level determined in (4.1) is correct, given that the parameters ρ and θ are known, which is not true in practice. So this section will consider the effect of estimating the parameters.

Only $\boldsymbol{\theta}$ unknown

First, only θ is considered to be unknown, so ρ and thus the coefficient of variation ($\nu = \rho^{-1/2}$) are assumed to be known; the last part of this subsection considers the situation that ρ is unknown as well. One possible estimator for θ is derived from the relation $\mathbb{E}[D] = \rho\theta$, leading to $\hat{\theta} = {}^{\bar{d}_t}/_{\rho} = \bar{d}_t\nu^2$, where $\bar{d}_t = \frac{1}{t}\sum_{i=1}^{t} d_i$ is the sample mean. The estimated order-up-to level in this case is then

$$\widehat{S} = F^{-1}_{\widetilde{\rho},1}(\alpha)\nu^2\bar{d}_t. \tag{4.2}$$

Define

$$g_\alpha = F^{-1}_{\widetilde{\rho},1}(\alpha)\nu^2. \tag{4.3}$$

Note that g_α consists of non-random terms and that $\bar{d}_t \sim \Gamma(t\rho, \frac{\theta}{t})$. This results in $\widehat{S} \sim \Gamma(t\rho, \frac{g_\alpha\theta}{t})$. Now let us consider the fraction of replenishment cycles with backlogged demand when using the order-up-to level $\widehat{S}$:

$$\mathbb{P}\left(D_{1+L} > \widehat{S}\right) = \mathbb{P}\left(\tfrac{1}{\theta}D_{1+L} > \tfrac{1}{\theta}\widehat{S}\right) = \mathbb{P}\left(D^*_{1+L} > \widehat{S}^*\right).$$

Note that $D^*_{1+L} = \frac{1}{\theta}D_{1+L} \sim \Gamma(\widetilde{\rho}, 1)$ and $\widehat{S}^* = \frac{1}{\theta}\widehat{S} \sim \Gamma(t\rho, \frac{g_\alpha}{t})$, so θ does not play a role in the derivation, which is intuitively clear, since θ is a scale parameter. Moreover, assuming $\widetilde{\rho} \in \mathbb{N}$ and $t\rho \in \mathbb{N}$, $f_{\widetilde{\rho},\cdot}(\cdot)$ and $f_{t\rho,\cdot}(\cdot)$ are density functions of an Erlang distributed variable, which is used in the derivation below, after (4.4):

$$\begin{aligned}
\mathbb{P}\left(D^*_{1+L} > \widehat{S}^*\right) &= \int_0^\infty \mathbb{P}\left(D^*_{1+L} > s\right) f_{t\rho,\frac{t}{g_\alpha}}(s)ds \\
&= \int_0^\infty (1 - F_{\widetilde{\rho},1}(s)) f_{t\rho,\frac{t}{g_\alpha}}(s)ds \qquad (4.4)\\
&= \int_0^\infty \left(e^{-s}\sum_{i=0}^{\widetilde{\rho}-1}\frac{(s)^i}{i!}\right)\left(\frac{\left(s\frac{t}{g_\alpha}\right)^{t\rho-1} e^{-s\frac{t}{g_\alpha}}\frac{t}{g_\alpha}}{(t\rho-1)!}\right) ds \\
&= \sum_{i=0}^{\widetilde{\rho}-1}\int_0^\infty e^{-s}e^{-s\frac{t}{g_\alpha}}\frac{s^i s^{t\rho-1}\left(\frac{t}{g_\alpha}\right)^{t\rho}}{i!(t\rho-1)!}ds \\
&= \sum_{i=0}^{\widetilde{\rho}-1}\frac{(t\rho-1+i)!}{\left(\frac{t}{g_\alpha}\right)^i(t\rho-1)!i!}\int_0^\infty e^{-s} f_{t\rho+i,\frac{t}{g_\alpha}}(s)ds \qquad (4.5)\\
&= \sum_{i=0}^{\widetilde{\rho}-1}\frac{(t\rho-1+i)!}{\left(\frac{t}{g_\alpha}\right)^i(t\rho-1)!i!}\left(\frac{\frac{t}{g_\alpha}}{\frac{t}{g_\alpha}+1}\right)^{t\rho+i} \\
&= \left(\frac{t}{t+g_\alpha}\right)^{t\rho}\sum_{i=0}^{\widetilde{\rho}-1}\binom{t\rho-1+i}{i}\left(\frac{g_\alpha}{t+g_\alpha}\right)^i. \qquad (4.6)
\end{aligned}$$

Note that the integral part of (4.5) is the Laplace-Stieltjes transform of an Erlang-$(t\rho + i)$ distribution. Now consider the special case of exponentially distributed demand ($\rho = 1$) and zero lead time. Then $g_\alpha = -\ln(1-\alpha)$ and (4.6) simplifies to

$$\mathbb{P}\left(D^*_{1+L} > \widehat{S}^*\right) = \left(\frac{t}{t+g_\alpha}\right)^t = \left(\frac{t}{t-\ln(1-\alpha)}\right)^t.$$

Now we prove that the desired service level is not reached, in other words that $(\frac{t}{t-\ln(1-\alpha)})^t > 1-\alpha$. The left hand side of this equation is a function of t and α: $h(t,\alpha) = \frac{t}{t-\ln(1-\alpha)})^t$. A graph of this function illustrates that indeed $h(t,\alpha)$ has $1-\alpha$ as its asymptote and reaches this asymptote from above; see Figure 4.1. Note that

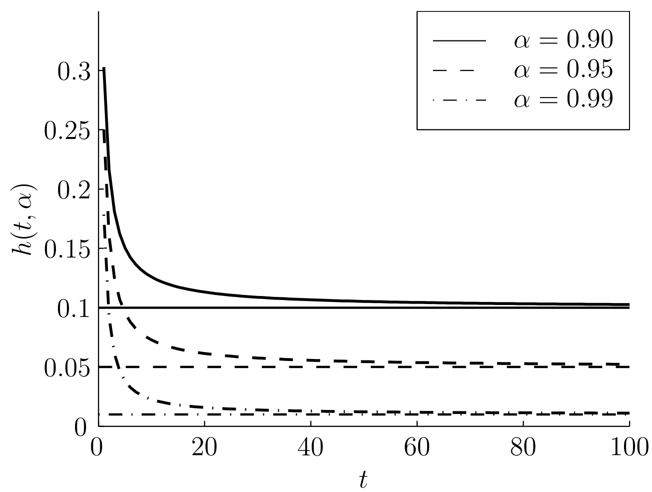


Figure 4.1: The function $h(t,\alpha)$ for three different values of α and its asymptote.

the difference between the attained stock out probability ($h(t,\alpha)$) and the desired stock out probability $(1-\alpha)$ is large when t is small, so it is not negligible. The conjecture $h(t,\alpha) > 1-\alpha$ can be rewritten as $(\frac{t+\ln((1-\alpha)^{-1})}{t})^t < (1-\alpha)^{-1}$; note that the left hand side is increasing in t. We finish our derivation as follows:

$$\left(1+\frac{\ln\left(\frac{1}{1-\alpha}\right)}{t}\right)^t < \lim_{t\to\infty}\left(1+\frac{\ln\left(\frac{1}{1-\alpha}\right)}{t}\right)^t = e^{\ln\left(\frac{1}{1-\alpha}\right)} = \frac{1}{1-\alpha}.$$

So in case of exponentially distributed demand and zero lead time, the desired service level will not be attained and the underperformance is larger if t is smaller. If we would replace α by $\alpha' = 1-\exp(t(1-(1-\alpha)^{-1/t}))$, where obviously $\alpha' > \alpha$, the desired service would be attained again. Unfortunately, we cannot find such analytical results if $L > 0$ or if $\rho \in \mathbb{N}/\{1\}$.

Now let us assume that $\widetilde{\rho} \in \mathbb{N}$ and $t\rho \in \mathbb{N}$, with $\rho \geq 2$. In this case we can use (4.6) to investigate the attained service level numerically. These calculations, not shown here, indicate that at low desired service levels the attained service level exceeds the desired one, while this reverses at higher values of α. If α approximates either 0 or 1, the attained service level approximates the desired one. This is easily explained: if $\alpha = 0$, $g_\alpha = 0$ and thus $\widehat{S} = 0$, so demand is not satisfied in any period. On the other hand, if $\alpha = 1$, $g_\alpha \to \infty$ and thus $\widehat{S} \to \infty$. In that case we can always satisfy demand. Further, there is one $\alpha \in (0, 1)$ such that the attained service equals the desired service level. These breakeven points are shown in Table 4.1 for several combinations of ρ, t and L. Table 4.1 clearly shows that the breakeven point gets

	$L = 0$			$L = 1$			$L = 3$		
ρ	$t = 2$	$t = 10$	$t = 20$	$t = 2$	$t = 10$	$t = 20$	$t = 2$	$t = 10$	$t = 20$
2	0.2499	0.2612	0.2627	0.3288	0.3474	0.3500	0.3688	0.3932	0.3971
6	0.3729	0.3817	0.3828	0.4055	0.4172	0.4189	0.4253	0.4400	0.4423
10	0.4038	0.4107	0.4116	0.4274	0.4366	0.4380	0.4423	0.4537	0.4556
20	0.4329	0.4380	0.4387	0.4489	0.4556	0.4565	0.4593	0.4674	0.4687
50	0.4579	0.4612	0.4616	0.4678	0.4720	0.4727	0.4743	0.4794	0.4803

Table 4.1: Values of α for which the non-stock-out probability using order-up-to level $\widehat{S}$ equals α (ρ known).

higher if ρ is larger, if t is larger or if L is larger. It also appears to converge to some value if ρ increases. In that case the coefficient of variation gets smaller and the gamma distribution looks more and more like a normal distribution. For this distribution it can easily be shown that the breakeven point is at $\alpha = 0.50$ (see (3.2) in Section 3.2).

In the remainder of this section and in Section 4.3 we will consider five different attained service levels. The notation of these service levels is denoted in Table 4.2.

$\hat{\alpha}_0(\hat{\beta}_0)$:	Attained service when θ unknown, but ρ known;
$\hat{\alpha}_1(\hat{\beta}_1)$:	Attained service when ρ and θ unknown and using $\alpha(\beta)$;
$\hat{\alpha}_2(\hat{\beta}_2)$:	Attained service when ρ and θ unknown and using $\alpha'(\beta')$;
$\hat{\alpha}_3(\hat{\beta}_3)$:	Attained service when using correction function $\hat{k}_\alpha(\hat{\rho}, t, \alpha, L)(\hat{k}_\beta(\hat{\rho}, t, \beta, L))$;
$\hat{\alpha}_4(\hat{\beta}_4)$:	Attained service when using correction function $\hat{k}_\alpha(\rho, t, \alpha, L)(\hat{k}_\beta(\rho, t, \beta, L))$.

Table 4.2: The attained service levels discussed in Sections 4.2 and 4.3 and their notation.

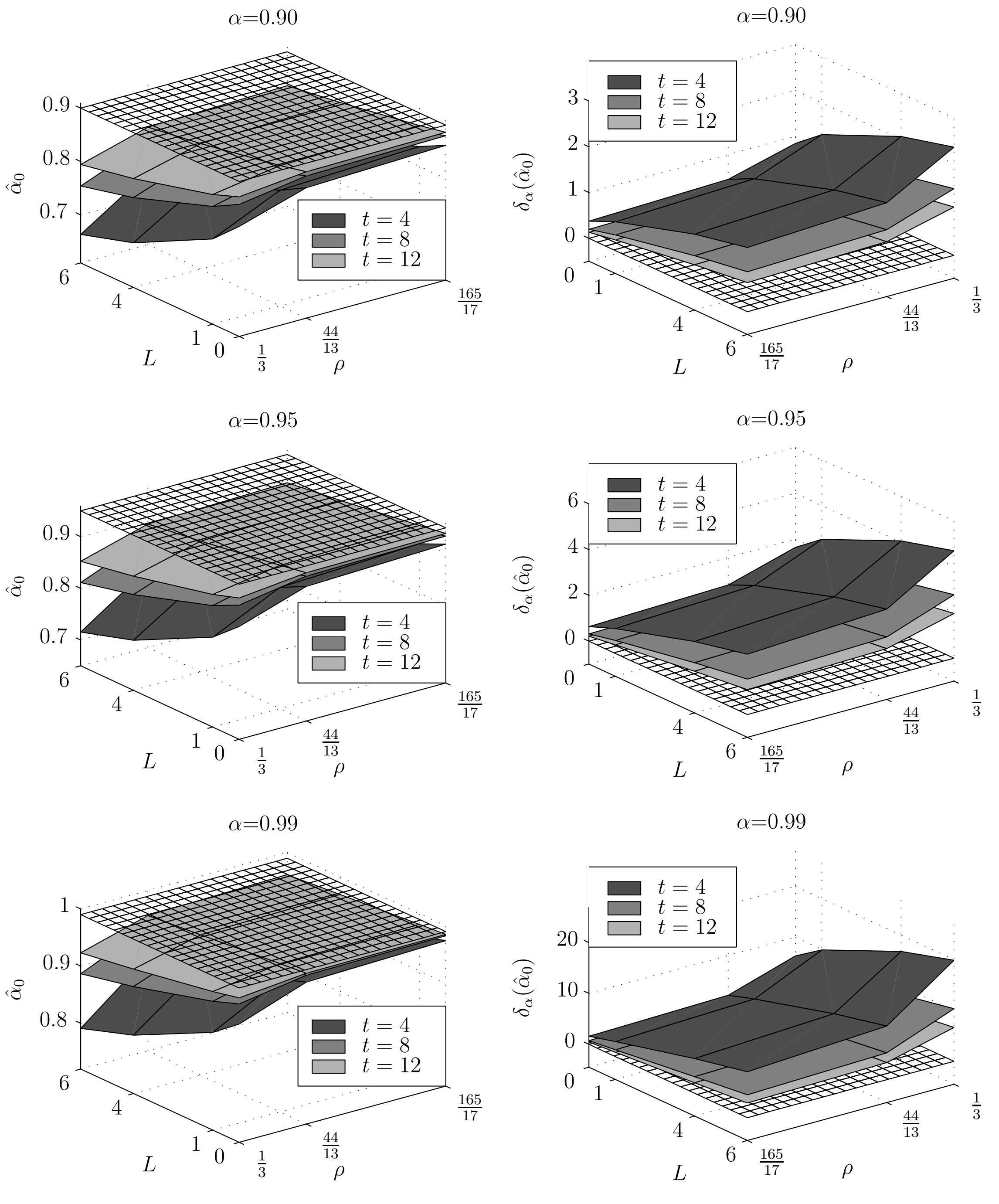


Figure 4.2: Attained service ($\hat{\alpha}_0$) and relative deviation ($\delta_\alpha(\hat{\alpha}_0)$) for non-integer but known values of ρ.

Now let us consider $\widetilde{\rho} > 0$ and $t\rho > 0$, not necessarily integer. In this case it is no longer possible to derive a closed-form expression, but we can numerically evaluate the integral in (4.4). The results of these evaluations are shown in Figure 4.2 for three non-integer values of ρ ($\rho \in \{\frac{1}{3}, \frac{44}{13}, \frac{165}{17}\}$). This figure show both the attained service and the relative deviation from the stock out probability; the latter is defined as $\delta_\alpha(\hat{\alpha}_0) = \frac{(1-\hat{\alpha}_0)-(1-\alpha)}{1-\alpha}$, where $\hat{\alpha}_0$ is the service level determined by evaluating (4.4). We have chosen this measure, since we think that the *perceived* customer service is mainly determined by stock out occurrences, so we choose to measure the performance relative to the desired probability of having a stock out $(1-\alpha)$. If $\delta_\alpha(\cdot)$ is positive, the attained service level is lower than the desired one and if $\delta_\alpha(\cdot)$ is negative, it is higher. Note that the z-axes of different desired service levels have different scales. Further, we have chosen to show the results only graphically; the corresponding numerical results listed in tables are in Appendix D.3. Only high values of α are considered, since these are used in practice.

The results in Figure 4.2 show that the earlier findings for $t\rho$ and $\widetilde{\rho}$ both integer also hold when these assumptions are relaxed: the desired service level is not reached. Furthermore we see that the underperformance is larger if ρ is smaller (ceteris paribus; c.p. for short), if t is smaller (c.p.), if L is larger (c.p.) and if α is larger (c.p.). If ρ is small, the coefficient of variation is large and hence demand is more variable, which implies that $\widehat{S}$ is more variable as well (cf. Chapter 3). Also if t is small, the estimator $\bar{d}_t$ is more variable and this also implies that $\widehat{S}$ is more variable. If L is large, we have to estimate demand for a longer period of time using the estimate of $\mathbb{E}[D]$ for one period. However, this estimate is multiplied by g_α and this factor is larger when L is larger (see (4.3)). So the error made by estimating $\mathbb{E}[D]$ is enlarged if L is larger and hence $\widehat{S}$ is more variable. The same line of reasoning applies to α being larger: in that case g_α is larger and the error made by estimating $\mathbb{E}[D]$ is enlarged, hence $\widehat{S}$ is more variable. So intuitively the relative deviation of the desired service level is larger if ρ and t are smaller and if L and α are larger; the numerical results confirm this intuition. The extreme deviations are listed in Table 4.3.

Desired service level	Minimum attained service ($\delta_\alpha(\hat{\alpha}_0)$)	Maximum attained service ($\delta_\alpha(\hat{\alpha}_0)$)	Mean attained service ($\delta_\alpha(\hat{\alpha}_0)$)
$\alpha = 0.90$	0.6635 (2.365)	0.8868 (0.132)	0.8213 (0.787)
$\alpha = 0.95$	0.7153 (4.694)	0.9393 (0.214)	0.8776 (1.448)
$\alpha = 0.99$	0.7910 (19.900)	0.9853 (0.470)	0.9411 (4.890)

Table 4.3: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ for non-integer but known values of ρ.

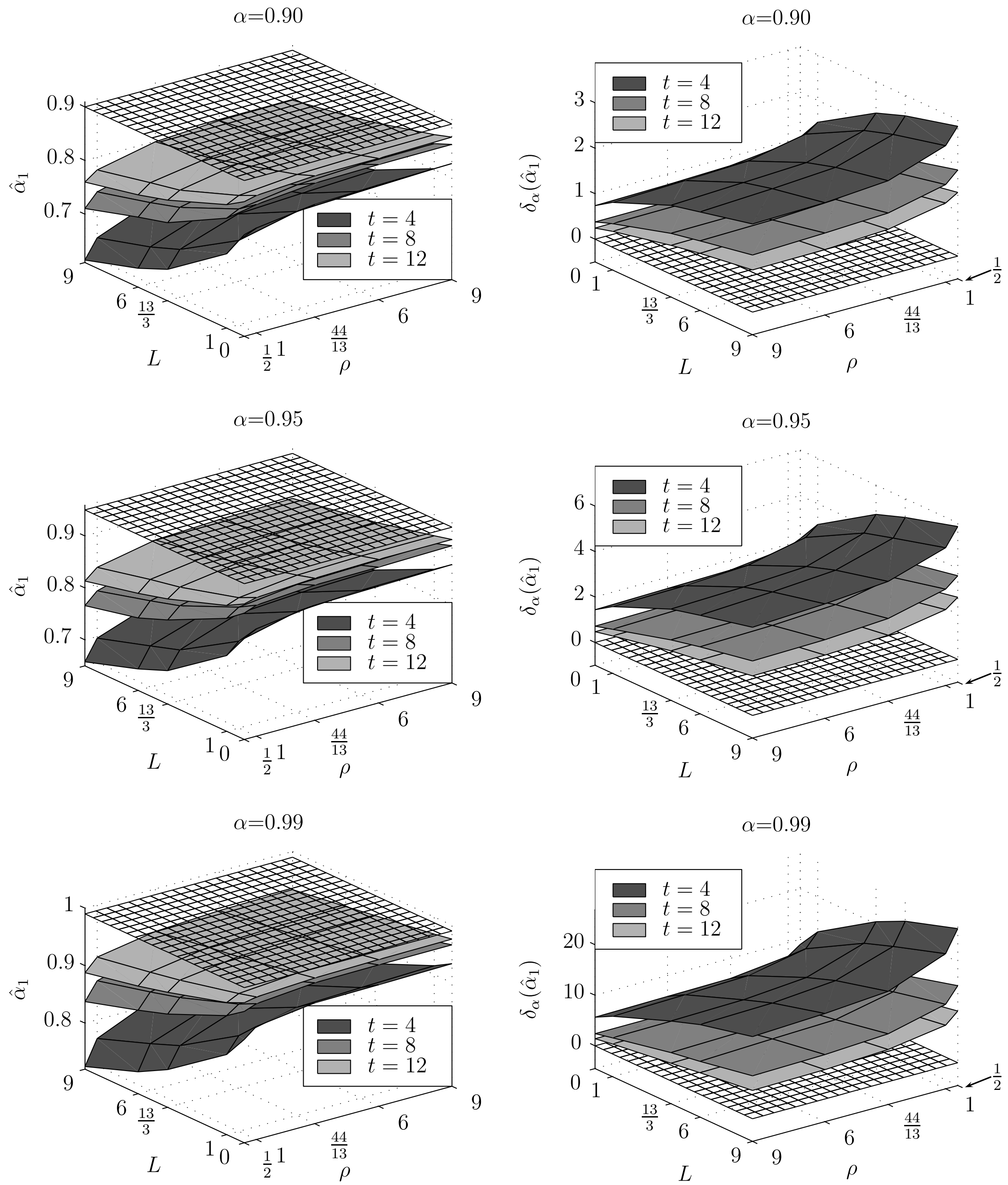


Figure 4.3: Simulated attained service ($\hat{\alpha}_1$) and relative deviation ($\delta_\alpha(\hat{\alpha}_1)$) when α is used (ρ unknown).

Both θ and ρ unknown

In the first part of this section ρ is assumed to be known. In the remainder of this subsection this assumption is relaxed, hence, an estimate of ρ is needed. The sample mean and sample variance $\left(s_t^2 = \frac{1}{t-1}\sum_{i=1}^{t}(d_i - \bar{d}_t)^2\right)$ are used to estimate ρ and θ: $\hat{\rho} = \bar{d}_t^2/s_t^2$ and $\hat{\theta} = s_t^2/\bar{d}_t$. Using these in the estimate of the order-up-to level results in $\widehat{S} = \hat{\theta}F^{-1}_{(1+L)\hat{\rho},1}(\alpha)$ and this is no longer gamma distributed. So derivations become intractable and hence simulation is used to determine the attained service level $\hat{\alpha}_1$ (see Appendix D.1). This simulation is restricted to high values of α, since values lower than 0.90 will not often be used in practice. The simulation has $n = 100,000$ replicates for each combination of ρ, t, α and L and its results are displayed in Figure 4.3. Remember that θ is a scale parameter and hence has no influence on the attained service level. Note that the scales of the z-axes of the graphs equal the scales of the z-axes of the graphs of Figure 4.2 with corresponding α-values for easy comparison. The same holds for the z-axes of Figures 4.4 and 4.6. Figure 4.3 clearly show that in all cases considered the desired service level is not reached. The underperformance again is larger if ρ is smaller (c.p.), if t is smaller (c.p.), if L is larger (c.p.) and if α is larger (c.p.). The extreme deviations and mean attained service level are displayed in Table 4.4.

Desired service level	Minimum attained service ($\delta_\alpha(\hat{\alpha}_1)$)	Maximum attained service ($\delta_\alpha(\hat{\alpha}_1)$)	Mean attained service ($\delta_\alpha(\hat{\alpha}_1)$)
$\alpha = 0.90$	0.6144 (2.856)	0.8847 (0.153)	0.8089 (0.911)
$\alpha = 0.95$	0.6575 (5.850)	0.9366 (0.268)	0.8624 (1.752)
$\alpha = 0.99$	0.7244 (26.560)	0.9819 (0.810)	0.9253 (6.470)

Table 4.4: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ when α is used (ρ unknown).

A first improvement

In the case of an exponential distribution ($\rho = 1$) and zero lead time, using α' instead of α ascertains that the desired service level is met. Using α' while ρ is unknown and $L \geq 0$ will probably not lead to meeting the desired service level, but since $\alpha' > \alpha$ it certainly increases the attained service level, denoted by $\hat{\alpha}_2$. The results of replacing α by α' are shown in Figure 4.4; Table 4.5 displays the extreme devi-

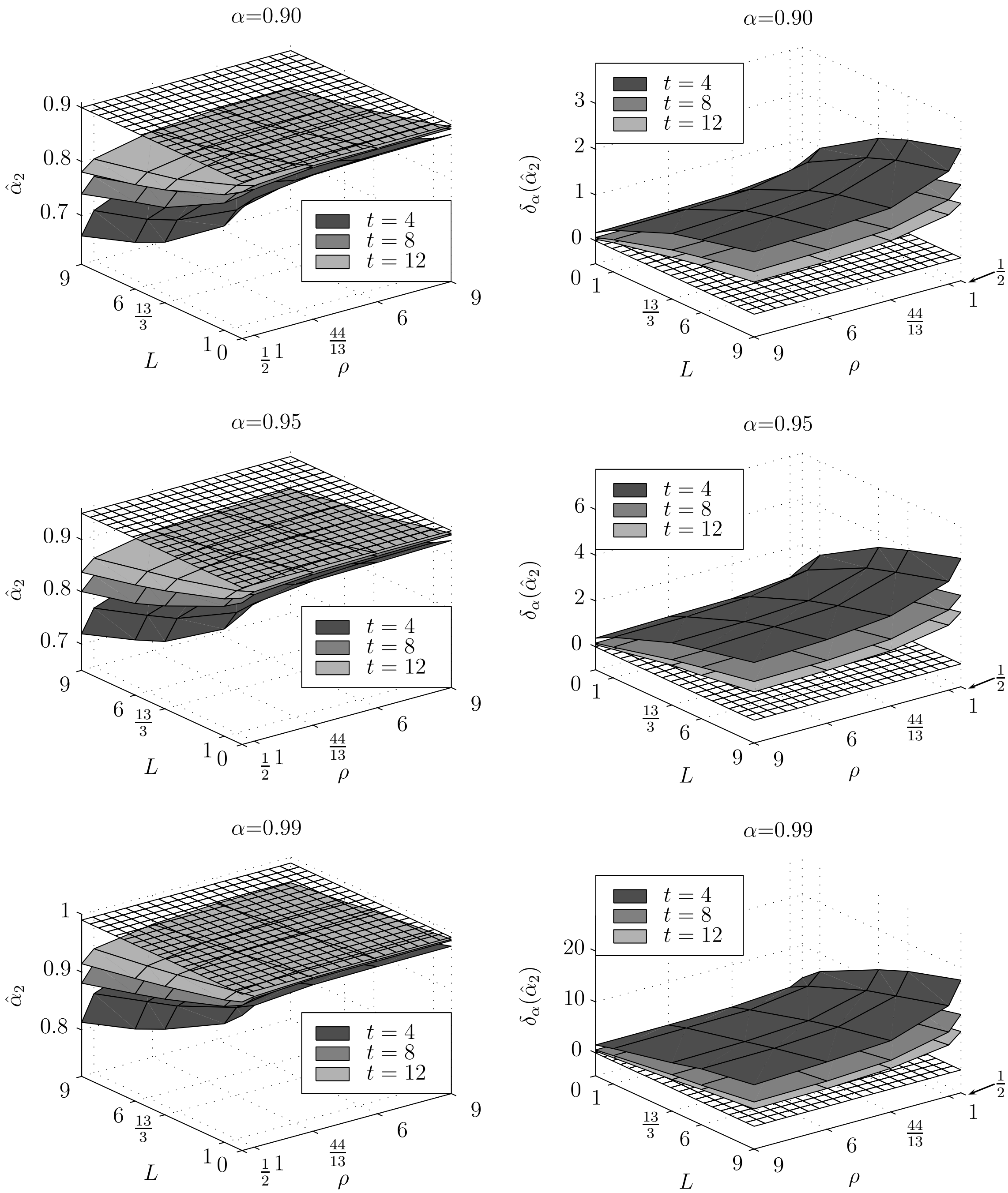


Figure 4.4: Simulated attained service ($\hat{\alpha}_2$) and relative deviation ($\delta_\alpha(\hat{\alpha}_2)$) when α' is used (ρ unknown).

ations. Figure 4.4 indeed shows that the performance improves significantly compared to using α; improvements range from almost 18% up to almost 80%. These relative improvements, denoted by $\mathcal{I}_\alpha(\hat{\alpha}_i, \hat{\alpha}_j)$, are measured by considering the percentage

Desired service level	Minimum attained service ($\delta_\alpha(\hat{\alpha}_2)$)	Maximum attained service ($\delta_\alpha(\hat{\alpha}_2)$)	Mean attained service ($\delta_\alpha(\hat{\alpha}_2)$)
$\alpha = 0.90$	0.6624 (2.376)	0.8975 (0.025)	0.8382 (0.618)
$\alpha = 0.95$	0.7197 (4.606)	0.9478 (0.044)	0.8933 (1.134)
$\alpha = 0.99$	0.8134 (17.660)	0.9881 (0.190)	0.9546 (3.540)

Table 4.5: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ when α' is used (ρ unknown).

change in the attained stock out probability. Let $\hat{\alpha}_i$ denote the attained service level using the order-up-to level determined by method i, here $\hat{\alpha}_1$, while $\hat{\alpha}_j$ denotes the attained service level using method j, $\hat{\alpha}_2$. The improvements $\mathcal{I}_\alpha(\hat{\alpha}_i, \hat{\alpha}_j)$ are calculated using

$$\mathcal{I}_\alpha(\hat{\alpha}_i, \hat{\alpha}_j) = \frac{\delta_\alpha(\hat{\alpha}_i) - \delta_\alpha(\hat{\alpha}_j)}{\delta_\alpha(\hat{\alpha}_i)} \cdot 100\% = \frac{\hat{\alpha}_j - \hat{\alpha}_i}{\alpha - \hat{\alpha}_i} \cdot 100\%.$$

So if $\mathcal{I}_\alpha(\hat{\alpha}_i, \hat{\alpha}_j)$ is between 0 and 100%, $\hat{\alpha}_j$ is closer to the desired service level than $\hat{\alpha}_i$. If it is negative, this reverses, i.e., the $\hat{\alpha}_i$ is closer to the desired service level than $\hat{\alpha}_j$. If $\mathcal{I}_\alpha(\hat{\alpha}_i, \hat{\alpha}_j)$ is larger than 100%, using method j instead of i results in overperformance, if under i there was underperformance and vice versa. If it is smaller than 200%, $\hat{\alpha}_j$ is closer to the desired service level compared to $\hat{\alpha}_i$ and this reverses if the improvement is larger than 200%.

Four cases are considered in detail; see Table 4.6. In the first case the attained

α	ρ	t	L	$\hat{\alpha}_1$ ($\delta_\alpha(\hat{\alpha}_1)$)	$\hat{\alpha}_2$ ($\delta_\alpha(\hat{\alpha}_2)$)	$\mathcal{I}_\alpha(\hat{\alpha}_1, \hat{\alpha}_2)$
0.90	$\frac{44}{13}$	8	$4\frac{1}{3}$	0.8016 (0.984)	0.8350 (0.650)	33.94%
0.95	9	12	1	0.9178 (0.644)	0.9375 (0.250)	61.18%
0.95	6	12	0	0.9262 (0.464)	0.9449 (0.102)	78.57%
0.99	$\frac{1}{2}$	4	6	0.7579 (23.210)	0.8445 (14.550)	37.31%

Table 4.6: Examples of improvement of attained service using α' instead of α.

service level increases from 0.8016 ($\delta_\alpha(\hat{\alpha}_1) = 0.984$) to 0.8350 ($\delta_\alpha(\hat{\alpha}_2) = 0.650$), which is an improvement of $(0.8350 - 0.8016)/(0.90 - 0.8016) \cdot 100\% = 33.94\%$. Using the relative deviations ($\delta_\alpha(\cdot)$), the same improvement is found: $(0.984 - 0.650)/0.984 \cdot 100\% = 33.94\%$. In the second case one can see that, although the attained service level is already pretty close to the desired service level (compared to the first and fourth case), a large improvement is possible using α' instead of α. This leads to almost reaching the desired service level. In the third case the desired service level is reached even closer. In the fourth case the attained service level is a lot closer to the desired service level, but there is still a large underperformance. This case has the most difficult parameter setting: both ρ and t are small while both α and L are large.

4.2.2 Determining the correction

As seen in Subsection 4.2.1 the desired service level is not met when using estimates in the determination of the order-up-to level. The attained service level can be improved by using α' instead of α, but still the desired service level is not reached. Another idea could be to use (an estimate of) the variance of the forecast error instead of the variance of demand (cf. Chapter 3). We have tried using this correction, but unfortunately simulation indicates no *consistent* improvement and hence we decided to consider other methods to improve the attained service level.

In this section it is shown that the attained service level is further improved (compared to only using α' instead of α) by using a multiplicative correction. That is, the estimated order-up-to level is multiplied by a certain factor that depends on the value of ρ, t, α and L (used parameter values are listed in Table 4.7). We also

Parameter	Values used in simulation
ρ	0.5, 1, 2, 4, 6, 8, 10
t	4, 6, 8, 10, 12, 20
α	0.90, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99
L	0, 0.5, 1, 3, 6

Table 4.7: Parameter values used for finding correction values.

tried to use a additive correction, as is used in Chapter 3, but this did not lead to consistent improvements. This might be intuitively explained by the functional form of the order-up-to level: if we assume gamma distributed demand, the order-up-to level is a multiplication of several terms, see (4.2). On the other hand, if demand is assumed to be normally distributed, the order-up-to level is a sum of two terms, see (3.1) and (3.4). Hence, a multiplicative correction might perform better for the gamma distribution, while an additive correction might be more suitable for the normal distribution.

Simulation is used to find the multiplication factor. The values of the corrections are found by first determining the order-up-to levels one would get while using $\hat{\rho}$, $\hat{\theta}$

and α'. Then the factor by which this order-up-to level should be multiplied in order to reach the desired service level is determined using bi-section. Figure 4.5 shows the corrections needed for different values of ρ, t, α and L.

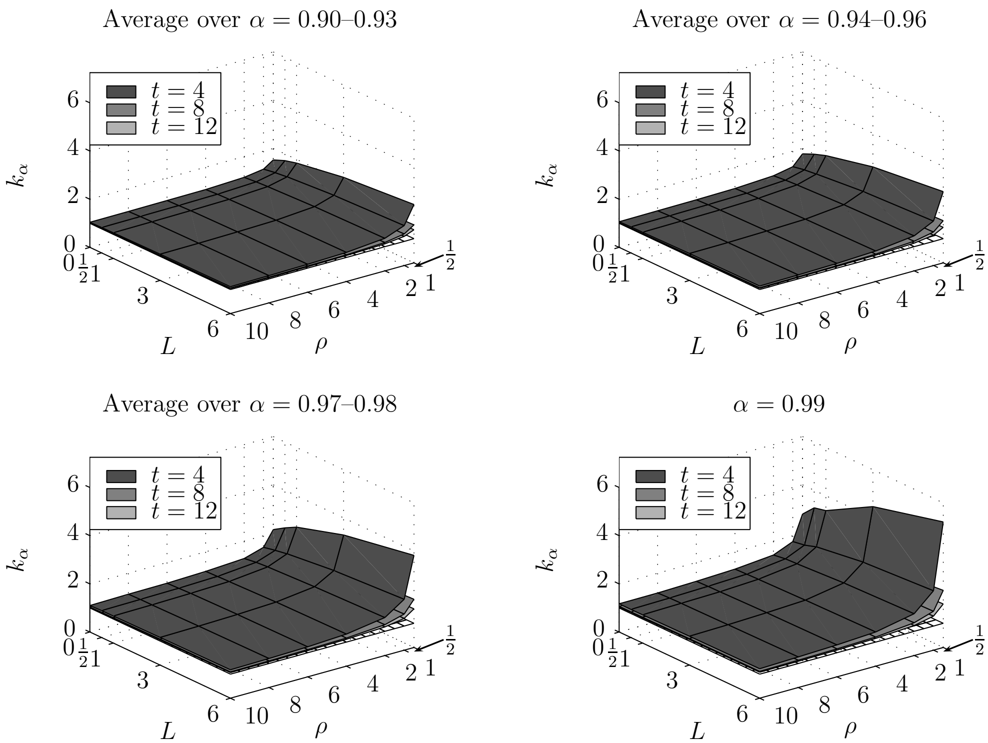


Figure 4.5: Corrections needed while using a P_1 criterion.

The corrections needed could be tabulated and then used to correct the order-up-to level, but a formula for the correction is easier to use. Nested regression (see Appendix B) is applied to find such a function. The natural logarithm of the correction is used as the dependent variable, since this transformation ensures that the correction increases more smoothly. The resulting function, $\hat{k}_\alpha(\rho, t, \alpha, L)$, is

$$\begin{aligned}
&\hat{k}_\alpha(\rho, t, \alpha, L) \\
&\quad = -0.0014 - 0.0988t^{-1.10} + (0.0005 + 0.0860t^{-1.80})a^{1.90} \\
&\quad + [0.0613 - 0.3845t^{-0.45} + (-0.0043 + 0.5375t^{-0.85})a^{0.85}]\rho^{-1.00} \\
&\quad + \big(-0.0282 + 0.0518t^{-0.15} + (0.0000 - 0.0231t^{-3.00})a^{2.75} \\
&\quad + [0.0703 - 0.0225t^{0.35} + (0.0044 + 0.1840t^{-1.45})a^{0.90}]\rho^{-0.75}\big)L^{0.55}.
\end{aligned} \tag{4.7}$$

Note that at the right hand side a is used, where $a = \ln((1-\alpha)^{-1})$ (α is the true desired service level, not the adapted value α'). The idea of using a instead of α

originates from the fact that $g_\alpha = a$ in case of exponentially distributed demand and zero lead time, hence a influences the order-up-to level directly. In short, (4.7) has been found as follows: we first choose only one dependent variable (L in case of (4.7)) and regress that on the logarithm of the correction needed, with different values for the power. We choose the power that results in the lowest sum of squared errors. Next we choose a second variable (ρ) and that is regressed on the coefficients found in the first regression. Also the third (a) and fourth (t) variable are treated in this manner; see Appendix B for a detailed description of the method. We choose to use this method, instead of, e.g., stepwise linear regression, because of the way the values of the powers are determined. Note that there are $4! = 24$ orderings for choosing the parameters; we have tried all orderings and have selected the best.

Using (4.7) on the parameter values listed in Table 4.7 results in an R^2 (coefficient of determination) of 0.9988 (adjusted $R^2 = 0.9987$), which is very high. However, using (4.7) implies that ρ is known, which is obviously not true in practice. We solve this by using the estimate for ρ, but that will be at the expense of a lower attained service. The size of the difference between the attained and desired service level is determined with help of simulation ($n = 100,000$) using (4.7). The order-up-to level in this simulation is thus determined by $\widehat{S} \cdot \exp(\hat{k}_\alpha(\hat{\rho}, t, \alpha, L))$. Figure 4.6 displays the resulting attained service levels ($\hat{\alpha}_3$) and the corresponding relative deviations ($\delta_\alpha(\hat{\alpha}_3)$); the extreme deviations are listed in Table 4.8. This simulation shows that

Desired service level	Minimum attained service ($\delta_\alpha(\hat{\alpha}_3)$)	Maximum attained service ($\delta_\alpha(\hat{\alpha}_3)$)	Mean attained service ($\delta_\alpha(\hat{\alpha}_3)$)
$\alpha = 0.90$	0.8387 (0.613)	0.9034 (-0.034)	0.8909 (0.091)
$\alpha = 0.95$	0.8899 (1.202)	0.9509 (-0.018)	0.9391 (0.218)
$\alpha = 0.99$	0.9479 (4.210)	0.9901 (-0.010)	0.9826 (0.740)

Table 4.8: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ using $\widehat{S} \cdot e^{\hat{k}_\alpha(\hat{\rho}, t, \alpha, L)}$.

indeed the desired service level is reached more closely; in case of ρ large, $\alpha = 0.90$ and $t = 12$ the desired service is even reached completely. Additional improvements for the remainder of the cases range from 60% to 99%. Total improvements for the cases in which the desired service level is not met range from 76% up to 99%.

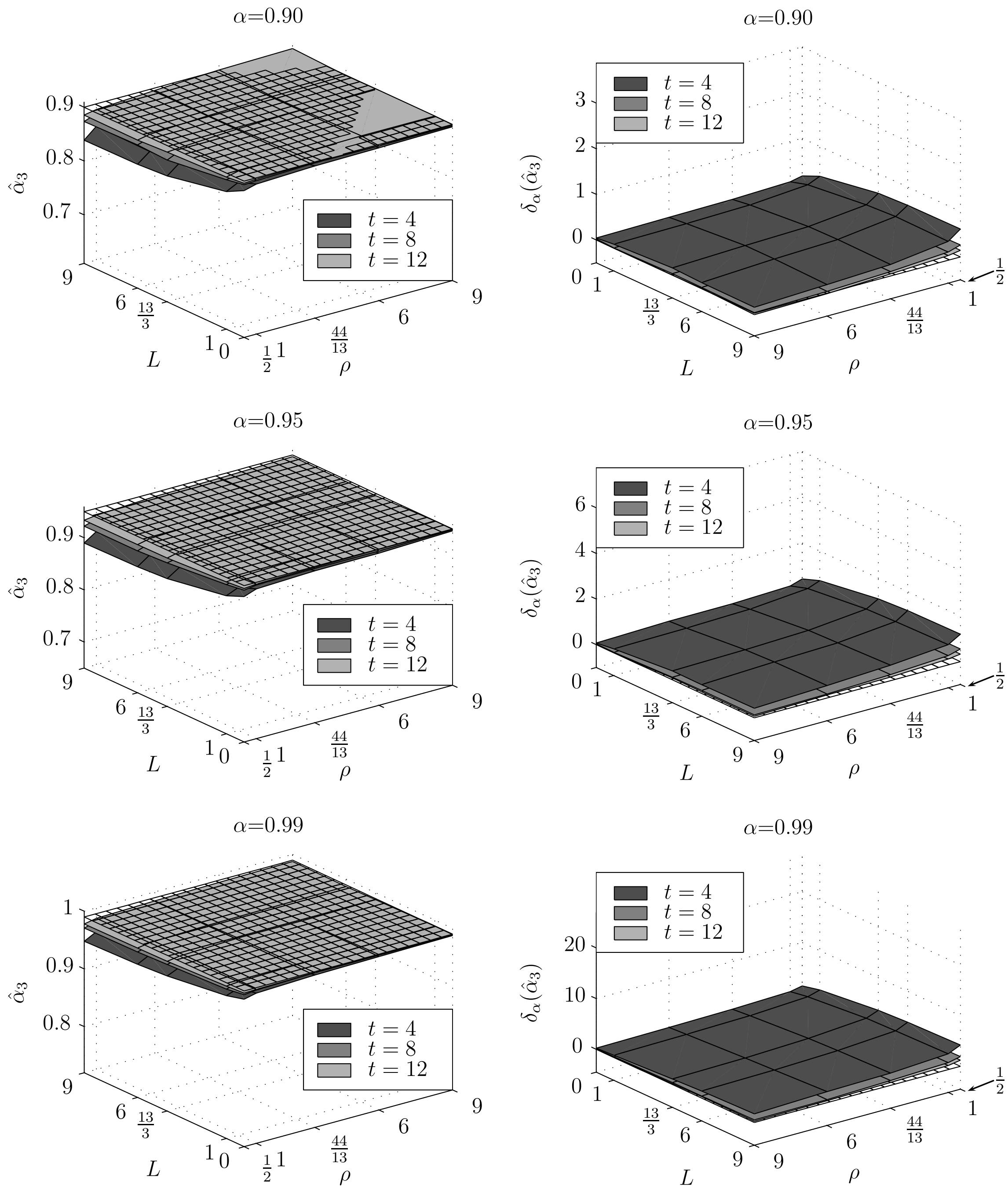


Figure 4.6: Simulated attained service ($\hat{\alpha}_3$) and relative deviation ($\delta_\alpha(\hat{\alpha}_3)$) using $\widehat{S} \cdot e^{\hat{k}_\alpha(\hat{\rho},t,\alpha,L)}$.

See Table 4.9 for the four cases that were also considered in Table 4.6. In the first case the attained service level is improved a lot and the desired service level is almost

α	ρ	t	L	$\hat{\alpha}_2$ $(\delta_\alpha(\hat{\alpha}_2))$	$\hat{\alpha}_3$ $(\delta_\alpha(\hat{\alpha}_3))$	$\mathcal{I}_\alpha(\hat{\alpha}_2, \hat{\alpha}_3)$	$\mathcal{I}_\alpha(\hat{\alpha}_1, \hat{\alpha}_3)$
0.90	$\frac{44}{13}$	8	$4\frac{1}{3}$	0.8350 (0.650)	0.8911 (0.089)	86.31%	90.96%
0.95	9	12	1	0.9375 (0.250)	0.9498 (0.004)	98.40%	99.38%
0.95	6	12	0	0.9449 (0.102)	0.9493 (0.014)	86.27%	97.06%
0.99	$\frac{1}{2}$	4	6	0.8445 (14.550)	0.9508 (3.920)	73.06%	83.11%

Table 4.9: Examples of improvement of attained service using the correction $\hat{k}_\alpha(\hat{\rho}, t, \alpha, L)$ instead of only α'.

reached. In the second and third case we can state that it actually is reached and in the fourth case we again see a large improvement upon the situation without using a correction, but unfortunately the underperformance is still quite large in this case.

The fact that the desired service level is not reached completely in almost all cases (see Figure 4.6), is due to using $\hat{\rho}$ instead of ρ. If this true value could be used, simulation shows (resulting in attained service levels $\hat{\alpha}_4$) that the desired service level is reached; the extreme deviations are denoted in Table 4.10.

Desired service level	Minimum attained service $(\delta_\alpha(\hat{\alpha}_4))$	Maximum attained service $(\delta_\alpha(\hat{\alpha}_4))$	Mean attained service $(\delta_\alpha(\hat{\alpha}_4))$
$\alpha = 0.90$	0.8972 (0.028)	0.9109 (-0.109)	0.9013 (-0.013)
$\alpha = 0.95$	0.9462 (0.076)	0.9556 (-0.112)	0.9499 (0.002)
$\alpha = 0.99$	0.9886 (0.140)	0.9919 (-0.190)	0.9898 (0.020)

Table 4.10: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ using $\widehat{S} \cdot e^{\hat{k}_\alpha(\rho,t,\alpha,L)}$.

We determined the estimates of the attained service in all our simulations in Section 4.2 using independent samples (each estimate was determined using different simulated demand periods). However, in practice one uses a rolling horizon. Our simulations have been repeated with a rolling horizon and the results are similar.

4.3 P_2 service level criterion

This section considers the P_2 service level criterion, so we need to find an order-up-to level S such that $\mathbb{E}\left[(D_{1+L} - S)^+\right] - \mathbb{E}\left[(D_L - S)^+\right] = (1 - \beta)\mathbb{E}\left[D\right]$. Note that shortages at the start of a replenishment cycle are included and if $L = 0$, the second expectation vanishes. When assuming that both parameters are known, it is not difficult to find the order-up-to level that satisfies this criterion. We have to find the value of S that satisfies the equality above. First note that the left hand side can be rewritten to

$$\begin{aligned}
&\mathbb{E}\left[(D_{1+L}-S)^+\right]-\mathbb{E}\left[(D_L-S)^+\right]\\
&\quad=\int_S^\infty (x-S)f_{\widetilde{\rho},\theta}(x)dx-\int_S^\infty (x-S)f_{L\rho,\theta}(x)dx=:\mathcal{L}_{\widetilde{\rho},\theta}(S)-\mathcal{L}_{L\rho,\theta}(S).
\end{aligned}$$

It is well known that $\mathcal{L}_{\rho,\theta}(y)=\int_y^\infty (x-y)f_{\rho,\theta}(x)dx=\rho\theta[1-F_{\rho+1,\theta}(y)]-y[1-F_{\rho,\theta}(y)]$. Note that there is no closed-form expression for the order-up-to level in general, hence we need to solve it numerically. However, if an exponential distribution is assumed, the order-up-to level using the P_2 criterion equals the order-up-to level using the P_1 criterion when $\alpha=\beta$ (see Appendix D.2). Hence, if $\rho=1$, $S=\theta F^{-1}_{1+L,1}(\beta)$.

4.3.1 Using estimates in determining the order-up-to level

The method described above can only be used when ρ and θ are known. This subsection discusses the effect of estimating ρ and θ on the attained P_2 service.

Only θ unknown

Let us first consider the case that only θ is unknown and $\rho=1$, hence we have exponentially distributed demand during the review period. Since in the case of exponentially distributed demand the theoretically correct order-up-to levels of the P_1 and the P_2 criterion are equal, the order-up-to level in case of the P_2 criterion can be estimated by $\widehat{S}=g_\beta\bar{d}_t$, where $g_\beta=F^{-1}_{1+L,1}(\beta)$. Then it is known that $\widehat{S}\sim\Gamma(t,\frac{g_\beta\theta}{t})$. The attained service, using again the fact that θ is a scale parameter and $\widehat{S}^*=\frac{1}{\theta}\widehat{S}$, is (using relation (D.2) in Appendix D.2)

$$\begin{aligned}
&\mathbb{E}\left[(D^*_{1+L}-\widehat{S}^*)^+\right]-\mathbb{E}\left[(D^*_L-\widehat{S}^*)^+\right]\\
&\quad=\mathbb{E}\left[\mathbb{E}\left[(D^*_{1+L}-\widehat{S}^*)^+\middle|\widehat{S}^*\right]-\mathbb{E}\left[(D^*_L-\widehat{S}^*)^+\middle|\widehat{S}^*\right]\right]\\
&\quad\overset{(D.2)}{=}\mathbb{E}\left[\mathbb{P}\left(D^*_{1+L}>\widehat{S}^*\middle|\widehat{S}^*\right)\right]=\mathbb{P}\left(D^*_{1+L}>\widehat{S}^*\right)\\
&\quad=\left(\frac{t}{t+g_\beta}\right)^t\sum_{i=0}^{L}\binom{t-1+i}{i}\left(\frac{g_\beta}{t+g_\beta}\right)^i\overset{L=0}{=}\left(\frac{t}{t+g_\beta}\right)^t>1-\beta.
\end{aligned}$$

Hence, in case of zero lead time the attained service always falls short of the desired one. In fact, as long as we are considering exponentially distributed demand during

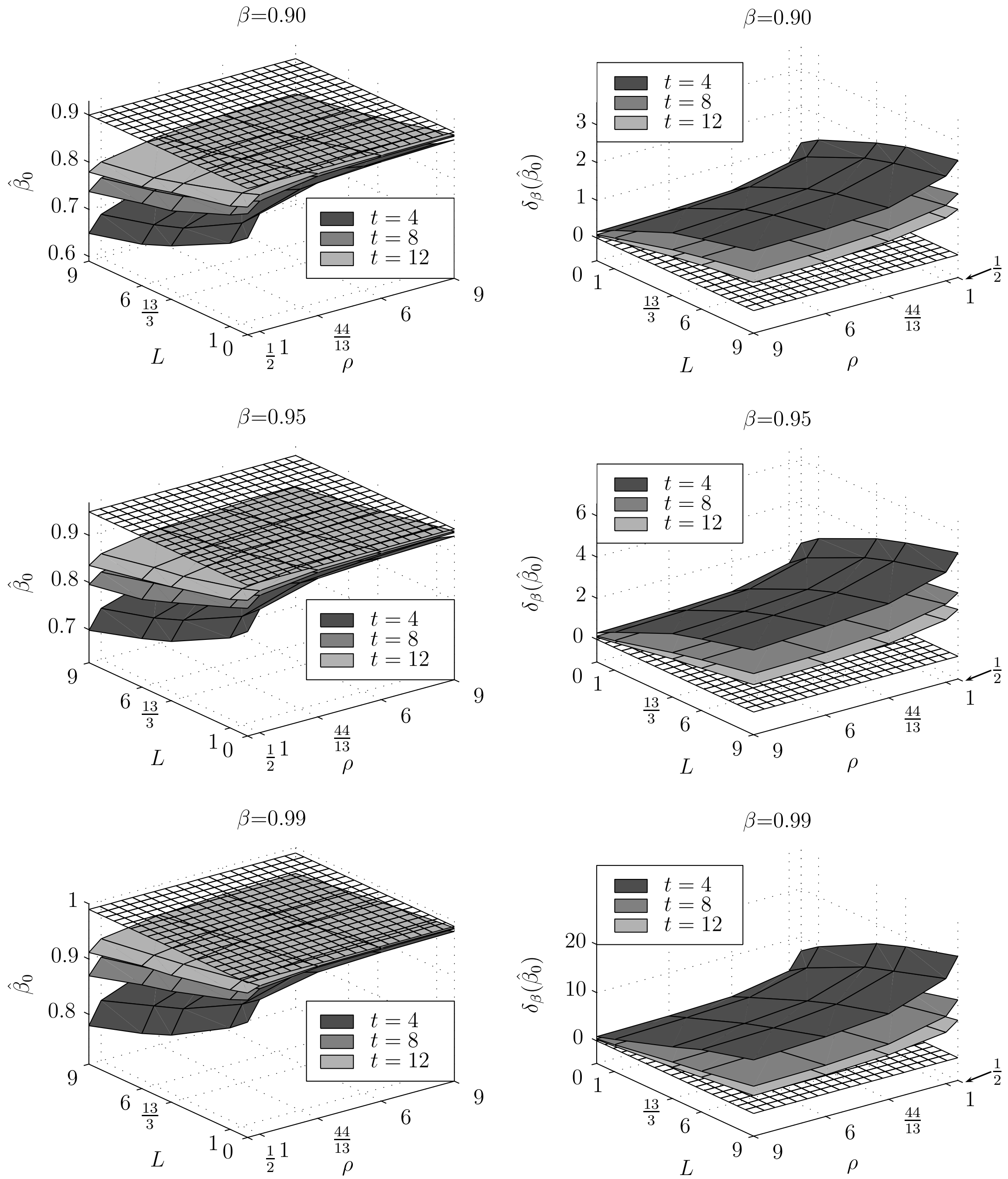


Figure 4.7: Simulated attained service ($\hat{\beta}_0$) and relative deviation ($\delta_\beta(\hat{\beta}_0)$) for non-integer but known values of ρ.

the review period, the results of the P_1 criterion also hold for the P_2 criterion. Unfortunately, if $\rho \neq 1$, no tractable results can be derived, due to the non-existence of a closed-form expression for $\widehat{S}$. Hence, simulation is used in this case (again with

$n = 100,000$ replicates) to obtain the performance of using an estimate for θ ($\hat{\theta} = \frac{\bar{d}_t}{\rho}$) in determining the order-up-to level. Note that ρ is still assumed to be known. The relative deviation from the desired fraction of backlogged demand, denoted by $\delta_\beta(\hat{\beta}_0) = \frac{(1-\hat{\beta}_0)-(1-\beta)}{1-\beta}$, where $\hat{\beta}_0$ is the service attained in simulation (see Appendix D.1), is shown in Figure 4.7; Table 4.11 displays the extreme deviations from the desired service level and the mean attained service levels. Note that the scales of

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_0)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_0)$)	Mean attained service ($\delta_\beta(\hat{\beta}_0)$)
$\beta = 0.90$	0.6491 (2.509)	0.8964 (0.036)	0.8303 (0.697)
$\beta = 0.95$	0.6987 (5.026)	0.9472 (0.056)	0.8863 (1.274)
$\beta = 0.99$	0.7794 (21.060)	0.9883 (0.170)	0.9487 (4.130)

Table 4.11: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ for non-integer but known values of ρ.

the graphs in Figure 4.7 of the z-axes vary across the different desired service levels. This simulation shows that the desired service level is not met when we have to use an estimate for θ instead of the true value. Also the underperformance is larger if ρ is smaller (c.p.), if t is smaller (c.p.), if L is larger (c.p.) and if β is larger (c.p.).

Both $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ unknown

If also ρ is assumed to be unknown, that parameter has to be estimated as well. The estimates for the parameters ρ and θ are $\hat{\rho} = \frac{\bar{d}_t^2}{s_t^2}$ and $\hat{\theta} = \frac{s_t^2}{\bar{d}_t}$, where $\bar{d}_t$ is the sample mean and s_t^2 the sample variance. The order-up-to level based on these estimates ($\widehat{S}$) is determined by solving (using binary search)

$$\mathcal{L}_{(1+L)\hat{\rho},1}(\widehat{S}^*) - \mathcal{L}_{L\hat{\rho},1}(\widehat{S}^*) = (1-\beta)\hat{\rho} \tag{4.8}$$

to get $\widehat{S}^*$ and then $\widehat{S} = \hat{\theta}\widehat{S}^*$. Note that the left hand side of (4.8) is decreasing in $\widehat{S}^*$, since it denotes the amount of backlogged demand, hence using binary search is appropriate. Next, simulation is used to estimate the performance of the order-up-to level determined in this way; $\hat{\beta}_1$ denotes the attained service. The results of this simulation, again based on $n = 100,000$ replicates, are shown in Figure 4.8. The z-axes in Figure 4.8 equal the z-axes in Figure 4.7 for easy comparison, just as the z-axes in Figures 4.9 and 4.11. Comparing Figure 4.8 to Figure 4.7 clearly shows that the underperformance is larger if both ρ and θ are assumed to be unknown.

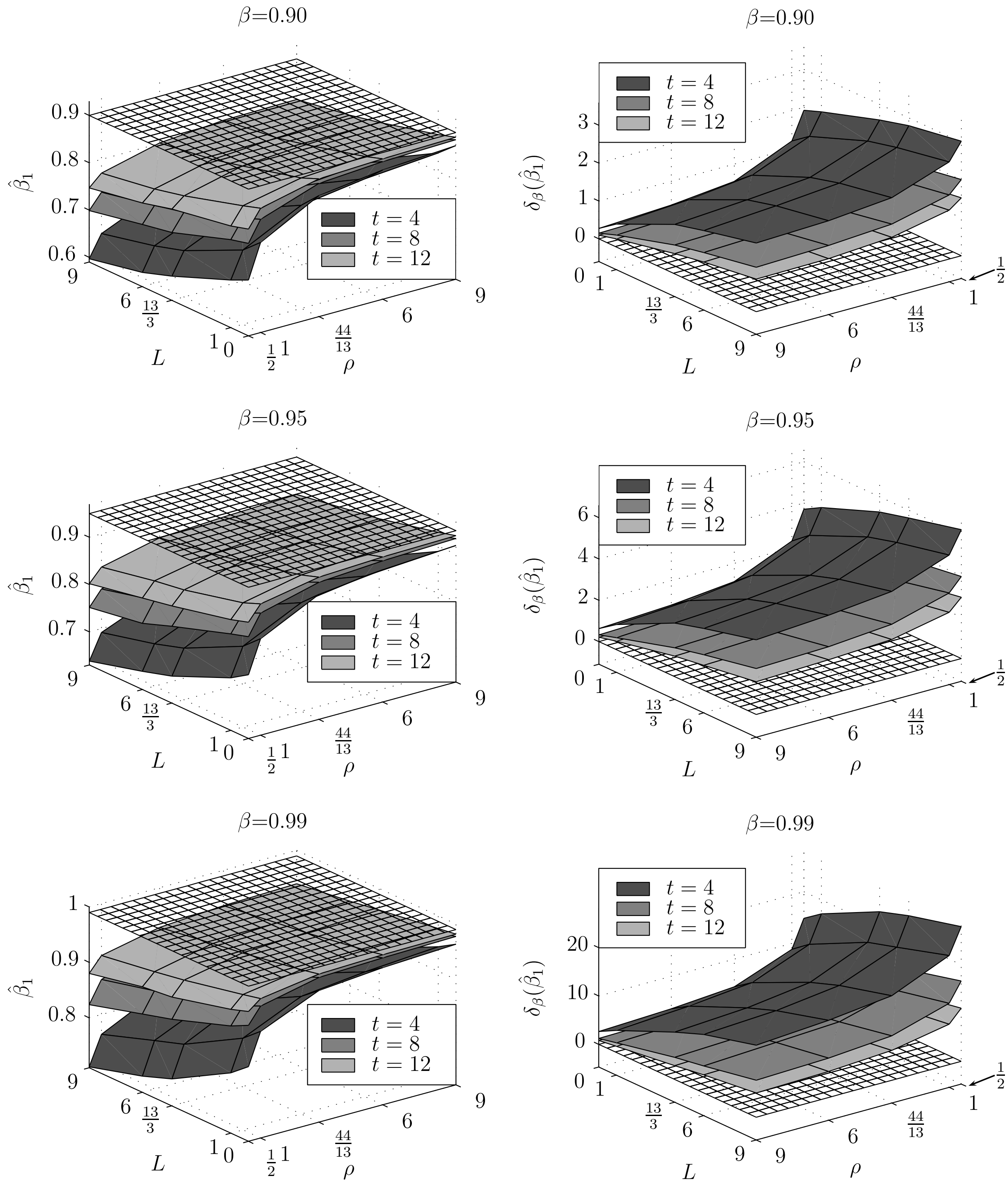


Figure 4.8: Simulated attained service ($\hat{\beta}_1$) and relative deviation ($\delta_\beta(\hat{\beta}_1)$) when β is used (ρ unknown).

Furthermore, we again see that the underperformance is larger when ρ is smaller (c.p.), when t is smaller (c.p.), when L is larger (c.p.) and when β is larger (c.p.). The extreme deviations from the desired service level are displayed in Table 4.12.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_1)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_1)$)	Mean attained service ($\delta_\beta(\hat{\beta}_1)$)
$\beta = 0.90$	0.5970 (3.030)	0.8938 (0.062)	0.8074 (0.926)
$\beta = 0.95$	0.6381 (6.238)	0.9434 (0.132)	0.8614 (1.772)
$\beta = 0.99$	0.7115 (27.850)	0.9853 (0.470)	0.9251 (6.490)

Table 4.12: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ when β is used (ρ unknown).

A first improvement

In case of exponential demand and zero lead time under the P_1 criterion the desired service level could be attained by using α' instead of α. Since the order-up-to levels under the P_1 and P_2 criterion are equal in case of exponential demand, the same holds here; i.e., if β is replaced by $\beta' = 1 - \exp(t(1-(1-\beta)^{-1/t}))$, the desired service level is met again, when lead time is zero. Of course, when demand is not exponentially distributed, this does not hold, but we can use this correction as a first improvement.

The attained service level ($\hat{\beta}_2$) and relative deviation ($\delta_\beta(\hat{\beta}_2)$) when using β' instead of β are shown in Figure 4.9; the extreme deviations in Table 4.13. Figure 4.9 clearly shows that, although the desired service is still not met (except for some

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_2)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_2)$)	Mean attained service ($\delta_\beta(\hat{\beta}_2)$)
$\beta = 0.90$	0.6472 (2.528)	0.9234 (-0.234)	0.8371 (0.629)
$\beta = 0.95$	0.7019 (4.962)	0.9616 (-0.232)	0.8927 (1.146)
$\beta = 0.99$	0.8026 (18.740)	0.9910 (-0.100)	0.9543 (3.570)

Table 4.13: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ when β' is used (ρ unknown).

cases where ρ large, $L = 0$ and t large), the attained service is significantly improved. In the cases where the desired service level is not met, improvements range from 16% to 99%; see Table 4.14 for the four cases of the previous examples. The improvements shown in this table are actually quite similar compared to the improvements for the P_1 criterion, except for the third case. In this case the desired service level is not only met, but even exceeded, which leads to an improvement of more than 100%, i.e., overperformance. Still, the attained service level is closer than before, since the absolute deviation now is 0.0038, while without using β' it was 0.0134.

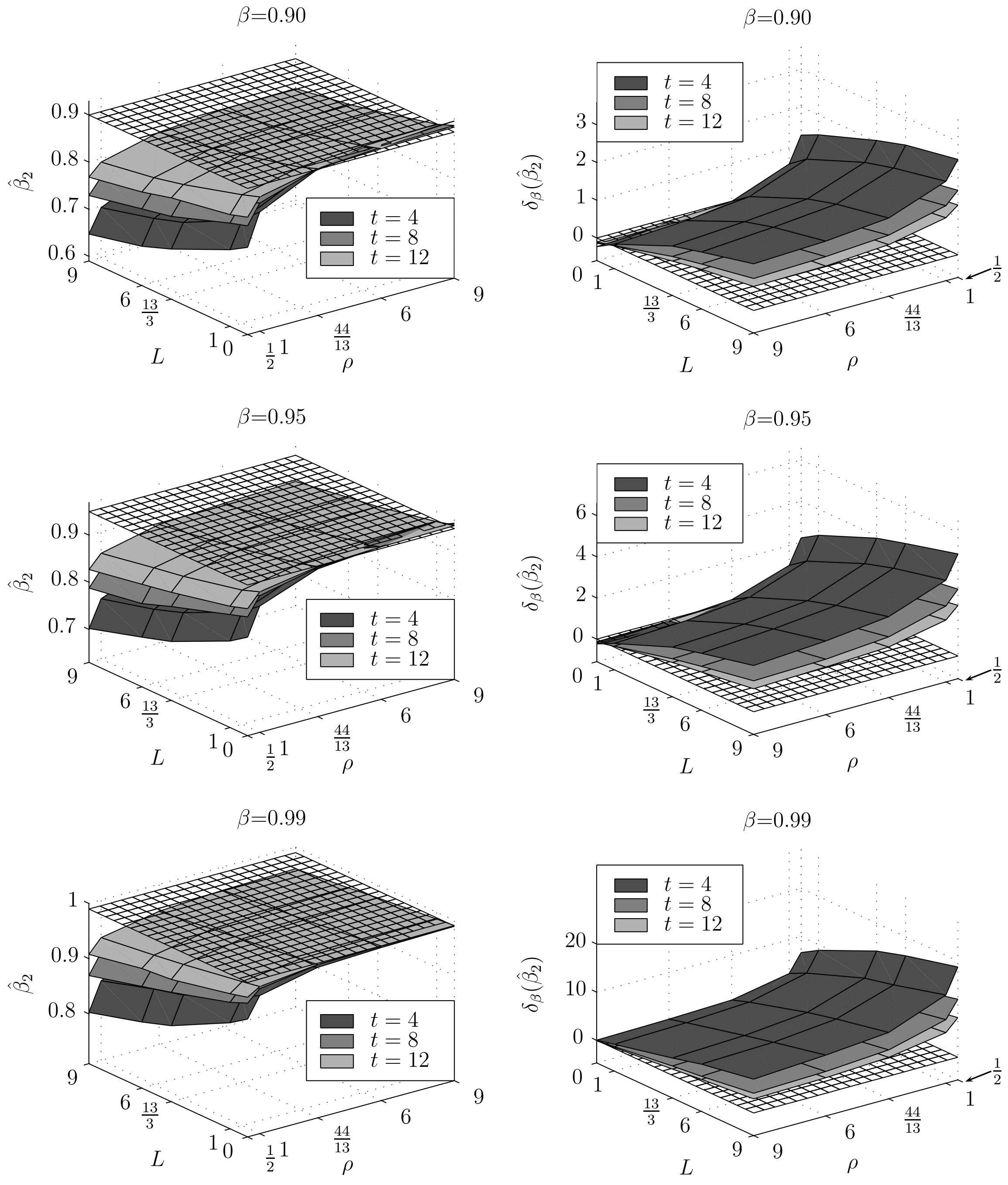


Figure 4.9: Simulated attained service ($\hat{\beta}_2$) and relative deviation ($\delta_\beta(\hat{\beta}_2)$) when β' is used (ρ unknown).

Therefore, the improvement is between 100% and 200%: the underperformance is changed to overperformance, but the absolute deviation from the desired service level is smaller. In the second case the desired service level is almost reached (closer

β	ρ	t	L	$\hat{\beta}_1$ ($\delta_\beta(\hat{\beta}_1)$)	$\hat{\beta}_2$ ($\delta_\beta(\hat{\beta}_2)$)	$\mathcal{I}_\beta(\hat{\beta}_1, \hat{\beta}_2)$
0.90	$\frac{44}{13}$	8	$4\frac{1}{3}$	0.8065 (0.935)	0.8391 (0.609)	34.87%
0.95	9	12	1	0.9294 (0.412)	0.9477 (0.046)	88.83%
0.95	6	12	0	0.9366 (0.268)	0.9538 (-0.076)	128.36%
0.99	$\frac{1}{2}$	4	6	0.7390 (25.100)	0.8295 (16.050)	36.06%

Table 4.14: Examples of improvement of attained service using β' instead of β.

compared to the P_1 criterion) and in the fourth case the attained service levels are a little below those of the P_1 criterion.

4.3.2 Determining the correction

Analogously to Section 4.2.2 we try to find a multiplicative correction in case a P_2 criterion is used. First the sizes of these corrections are determined with help of simulation: the order-up-to levels based on $\hat{\rho}$, $\hat{\lambda}$ and β' are calculated for different values of ρ, t, β and L (see Table 4.7, with $\alpha = \beta$). The corrections needed to attain the service level are then determined numerically and these corrections for different combinations of ρ, t, β and L are shown in Figure 4.10.

The same linear regression technique as outlined in Section 4.2.2 is used to find a function that estimates the logarithms of the corrections needed and the resulting function is

$$\begin{aligned}
&\hat{k}_\beta(\rho, t, \beta, L) \\
&\quad = -0.0154 - 1.0112t^{-1.25} + (-0.1363 + 0.2797t^{-0.20})\rho^{-1.45} \\
&\quad + [0.0034 + 0.4644t^{-1.15} + (0.0082 - 0.2634t^{-0.75})\rho^{-1.15}]L^{0.35} \\
&\quad + \big(-0.0014 + 1.2026t^{-2.90} + (0.0230 + 0.7037t^{-1.05})\rho^{-0.85} \\
&\quad + [0.0029 - 17.2361t^{-5.85} + (-0.0034 + 0.1449t^{-1.00})\rho^{-0.80}]L^{0.55}\big)b^{0.85},
\end{aligned} \tag{4.9}$$

where $b = \ln((1-\beta)^{-1})$. Using this function to estimate the corrections needed results in an R^2 of 0.9987 (adjusted $R^2 = 0.9987$), which is again very high. However, this function implies that one needs to know the true value of ρ, which is not known in practice. Hence, $\hat{\rho}$ is used to calculate the correction needed and simulation ($n = 100,000$) is performed to determine the effect of using $\widehat{S} \cdot \exp(\hat{k}_\beta(\hat{\rho}, t, \beta, L))$ to estimate the order-up-to level; the attained service level is denoted by $\hat{\beta}_3$. In this order-up-to level $\widehat{S}$ is determined by using the adapted desired service level β'. The correction is determined using the true desired service level β. Like in the case of the

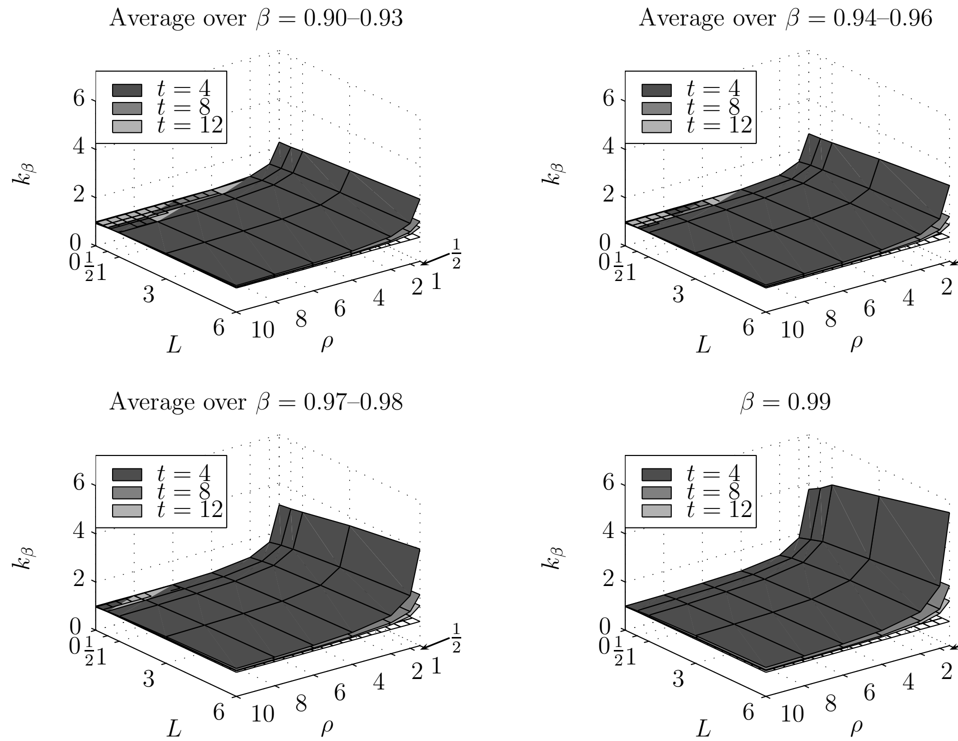


Figure 4.10: Corrections needed while using a P_2 criterion.

P_1 criterion, the coefficients are rounded to 10^{-4}. The results are shown in Figure 4.11; the extreme deviations are in Table 4.15. This simulation shows that the desired service level is reached more closely, but not reached completely yet; also in the few cases it was reached when not using this correction function ($\hat{\beta}_2$; see Figure 4.9) the service is not reached, since the correction needed is smaller than 1 for those cases. In general, additional improvements range from only a few percent or even a little decline, in the cases were the desired service level was (almost) met, to 94%. The total improvements range from 66% to 94%. Table 4.16 shows the four cases we have considered earlier.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_3)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_3)$)	Mean attained service ($\delta_\beta(\hat{\beta}_3)$)
$\beta = 0.90$	0.8231 (0.769)	0.9066 (-0.066)	0.8859 (0.141)
$\beta = 0.95$	0.8781 (1.438)	0.9541 (-0.082)	0.9360 (0.280)
$\beta = 0.99$	0.9447 (4.530)	0.9908 (-0.080)	0.9810 (0.900)

Table 4.15: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ using $\widehat{S} \cdot e^{\hat{k}_\beta(\hat{\rho},t,\beta,L)}$.

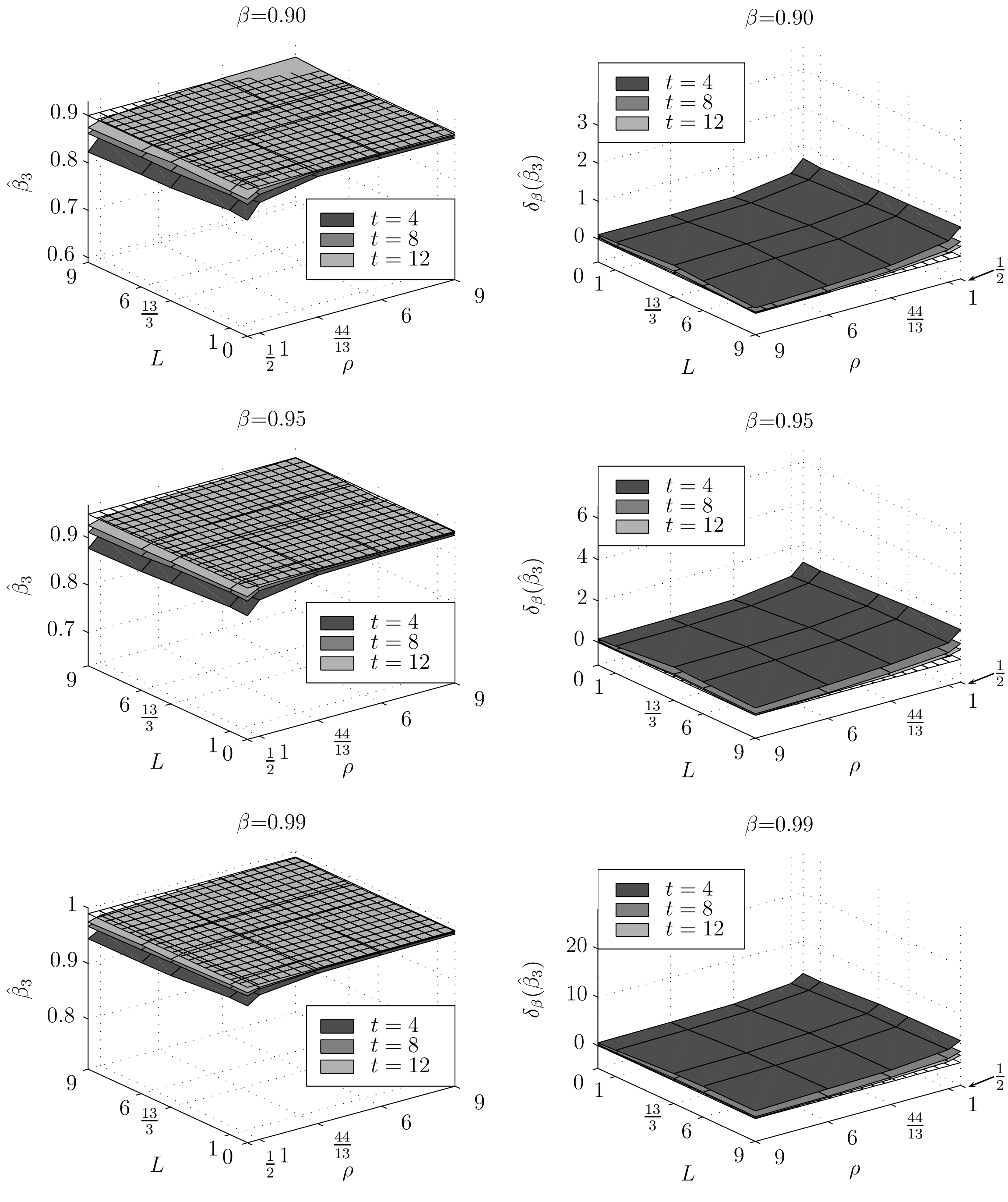


Figure 4.11: Simulated attained service ($\hat{\beta}_3$) and relative deviation ($\delta_\beta(\hat{\beta}_3)$) using $\widehat{S} \cdot e^{\hat{k}_\beta(\hat{\rho},t,\beta,L)}$.

In the first case the attained service level is improved upon considerably and the desired service level is almost reached. In the second case we only see a minor

β	ρ	t	L	$\hat{\beta}_2$ $(\delta_\beta(\hat{\beta}_2))$	$\hat{\beta}_3$ $(\delta_\beta(\hat{\beta}_3))$	$\mathcal{I}_\beta(\hat{\beta}_2, \hat{\beta}_3)$	$\mathcal{I}_\beta(\hat{\beta}_1, \hat{\beta}_3)$
0.90	$\frac{44}{13}$	8	$4\frac{1}{3}$	0.8391 (0.609)	0.8881 (0.119)	80.46%	87.27%
0.95	9	12	1	0.9477 (0.046)	0.9484 (0.032)	30.43%	92.23%
0.95	6	12	0	0.9538 (-0.076)	0.9486 (0.028)	136.84%	89.55%
0.99	$\frac{1}{2}$	4	6	0.8295 (16.050)	0.9459 (4.410)	72.52%	82.43%

Table 4.16: Examples of improvement of attained service using the correction $\hat{k}_\beta(\hat{\rho}, t, \beta, L)$ instead of only β'.

improvement, due to the fact that the attained service level was already very close to the desired one. The third case shows what is mentioned above, i.e., due to the fact that the correction needed is smaller than one, the attained service level declines a little. It is again below, but very close to the desired service level. Finally, in the fourth case the underperformance is still substantial, however, using the correction improves the attained service considerably.

Again, not reaching the desired service level completely is due to the fact that $\hat{\rho}$ is used instead of ρ. If the true value of ρ is used, simulation (resulting in attained service levels $\hat{\beta}_4$) shows that the desired service level is reached; the extreme deviations are denoted in Table 4.17.

Desired service level	Minimum attained service $(\delta_\beta(\hat{\beta}_4))$	Maximum attained service $(\delta_\beta(\hat{\beta}_4))$	Mean attained service $(\delta_\beta(\hat{\beta}_4))$
$\beta = 0.90$	0.8944 (0.056)	0.9130 (-0.130)	0.9003 (-0.003)
$\beta = 0.95$	0.9455 (0.090)	0.9577 (-0.154)	0.9501 (-0.002)
$\beta = 0.99$	0.9887 (0.130)	0.9920 (-0.200)	0.9900 (0.000)

Table 4.17: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ using $\widehat{S} \cdot e^{\hat{k}_\beta(\rho,t,\beta,L)}$.

Analogous to Section 4.2, we have used independent samples for the P_2-case as well. The simulations with a rolling horizon, results not shown, yield similar attained service levels.

4.4 Case study

The consultancy firm Involvation provided daily demand data for a period of 9 years and 4 months of the Dutch Ministry of Defence, which is one of their customers. The data is from a department that manages inventories of all kinds of not perishable products, ranging from screws and bolts to first aid equipment and spare parts for vehicles. We use these data to test the method for determining the order-up-to level developed in Sections 4.2 and 4.3. Since gamma demand is assumed, the demand should be continuous. However, with daily demand this is difficult to accomplish, so we aggregated the daily demand data to monthly demand data (resulting in 112 months). The review period R is set to 1 month and the lead time L is expressed in months. Still a lot of products have intermittent demand and therefore we selected products that faced positive monthly demand for at least 17 consecutive months; 2462 products satisfied this requirement. Of those 2462 products 602 had two or more non-overlapping periods with monthly demand occurrence for at least 17 consecutive months, resulting in 3153 demand streams to work with. The length of the demand streams range from 17 to 112 months, mostly (84%) less than 50 months. Next, we used a two-tailed Anderson-Darling test to test whether each of the 3153 demand streams are gamma distributed or not at a significance level of 5%. In 401 cases there is evidence to support that the demand is *not* gamma distributed; in the remaining 2752 cases there is not enough evidence. However, we do not discard those 401 cases from our simulation.

We have not identified which products could be nonstationary. However, the Dutch army is reducing in size and therefore less products are demanded. Hence, at least part of the demand streams exhibits a downward trend. Therefore, using a moving average to forecast demand is not illogical.

Since the shape parameter ρ is now determined by the demand data available, we only have 3 parameters left: the number of observations used ($t \in \{4, 8, 12\}$), the desired service level ($\alpha(\beta) \in \{0.90, 0.95, 0.99\}$) and the lead time ($L \in \{0, 1, 4\}$). For every demand stream we determined the order-up-to levels under the $P_1(P_2)$ criterion using $\alpha(\beta)$, $\alpha'(\beta')$ and the correction $\hat{k}_1(\hat{k}_2)$ for every combination of t, $\alpha(\beta)$ and L as follows. Since we need independent observations of the stock out occurrences (P_1) and the demand backlogged (P_2), the data streams were split up into parts, each containing $t + L + 1$ observations (therefore we selected products with at least 17 consecutive periods with demand: $12 + 4 + 1 = 17$). The first t periods are used to estimate ρ and θ and those estimates are used to find the order-up-to levels. Next, if $L \neq 0$, the demand during lead time (the $(t+1)$th up till the $(t + L)$th observation) is subtracted from the order-up-to level to get the net inventory at the start of the replenishment cycle. Finally the demand in the replenishment cycle (the $(t+L+1)$th observation) is subtracted from the net inventory at the start of the replenishment cycle, which results in the net inventory at the end of the replenishment cycle. See Figure 4.12 for an example where we have a demand stream containing 20 periods, $t = 4$ and $L = 1$. The demand during the replenishment cycle is denoted by D_{RC},

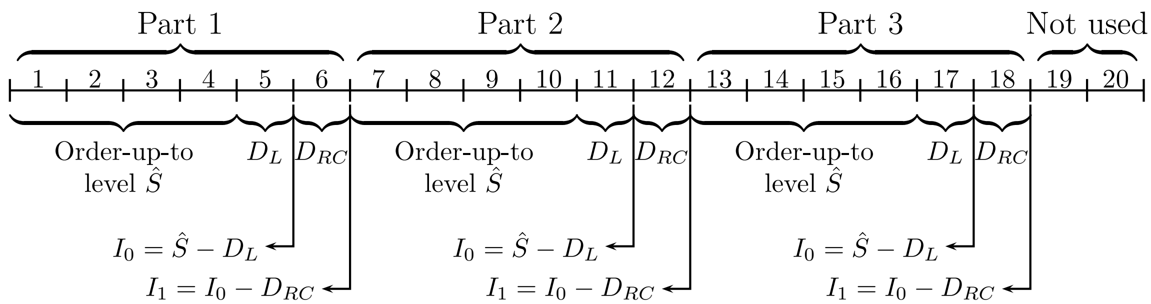


Figure 4.12: Example of obtaining net inventory in the case study with $t = 4$ and $L = 1$.

I_0 denotes the net inventory at the start of the replenishment cycle, and I_1 the net inventory at the end of the replenishment cycle. We split this demand stream in three parts, each containing six demand observations. We have two observations left which are not used. The order-up-to level used in part one is determined by the demand observations in months 1 up to and including 4. The net inventory at the start of the first replenishment cycle (at the start of month 6) is determined by subtracting the demand of month 5 from the order-up-to level. Finally, the net inventory at the end of the first replenishment cycle is determined by subtracting the demand of month 6 from I_0. In the same way the net inventories of parts two and three are determined.

For P_1 service we need the number of replenishment cycles with a stock out occurrence, so the number of times a negative net inventory at the end of the replenishment cycle occurs is counted. For P_2 service we need the amount of demand that cannot be satisfied immediately, so the backlog in all replenishment cycles is summed. The backlog during a replenishment cycle is determined by $(-I_1)^+ - (-I_0)^+$.

The net inventory at the start and at the end of the replenishment cycle are used to determine the attained service level, denoted by $\hat{\alpha}_C$ for the cycle service and $\hat{\beta}_C$ in case of the fill rate.

Figure 4.13 displays the resulting aggregated attained service for the P_1 criterion; Table 4.18 displays the extreme deviations while using α, α' and $\hat{k}_\alpha(\hat{\rho}, t, \alpha, L)$. If we consider the 3153 demand streams individually, we find that some of them have an attained service that is higher than the desired one, while there are also demand streams

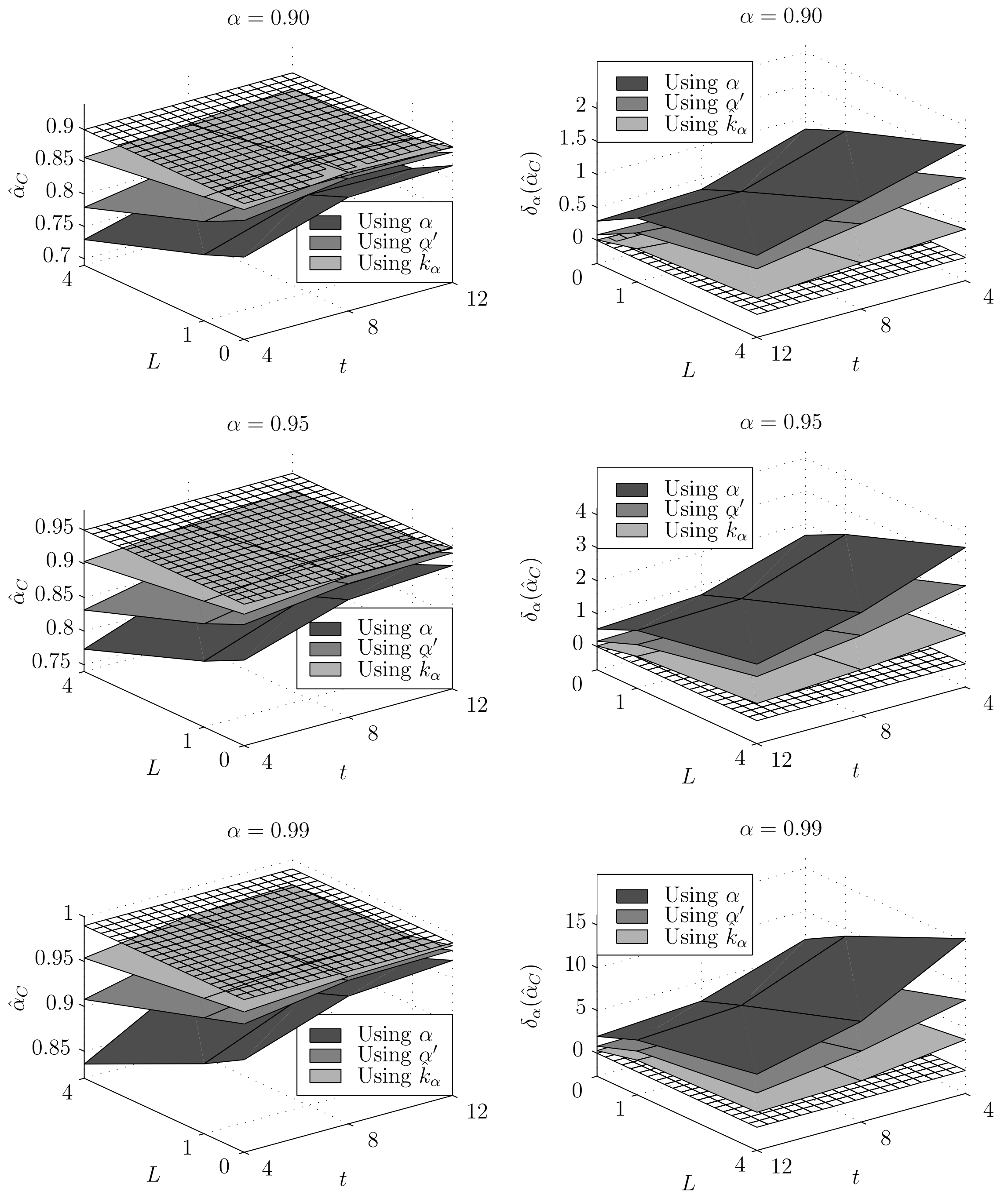


Figure 4.13: Attained service ($\hat{\alpha}_C$) and relative deviation ($\delta_\alpha(\hat{\alpha}_C)$) using the P_1 criterion in the case study.

with an attained service lower than the desired one. If we consider the aggregated service of all demand streams, we see that the desired service is not reached. One can

Desired service level	Minimum attained service ($\delta_\alpha(\hat{\alpha}_C)$)	Maximum attained service ($\delta_\alpha(\hat{\alpha}_C)$)	Mean attained service ($\delta_\alpha(\hat{\alpha}_C)$)
$\alpha = 0.90$			
Using α	0.7297 (1.703)	0.8709 (0.291)	0.8153 (0.847)
Using α'	0.7796 (1.204)	0.8950 (0.050)	0.8504 (0.496)
Using $\hat{k}_\alpha$	0.8569 (0.431)	0.9044 (-0.044)	0.8839 (0.161)
$\alpha = 0.95$			
Using α	0.7733 (3.534)	0.9232 (0.536)	0.8665 (1.670)
Using α'	0.8315 (2.370)	0.9417 (0.166)	0.9012 (0.976)
Using $\hat{k}_\alpha$	0.9033 (0.934)	0.9493 (0.014)	0.9312 (0.376)
$\alpha = 0.99$			
Using α	0.8357 (15.430)	0.9704 (1.960)	0.9239 (6.610)
Using α'	0.9077 (8.230)	0.9818 (0.820)	0.9553 (3.470)
Using $\hat{k}_\alpha$	0.9539 (3.610)	0.9870 (0.300)	0.9744 (1.560)

Table 4.18: Extreme deviations from desired service level for $\alpha \in \{0.90, 0.95, 0.99\}$ in the case study.

clearly see in Table 4.18 that using α' instead of α results in less underperformance; improvements range from 22% to 85%. It can be verified (results not shown here), using a t-test with a significance level of 5%, that the improvement is significant, except for two cases ($t = 12$, $L = 0$ and $\alpha = 0.90, 0.95$). Using the correction instead of just α' again results in less underperformance; additional improvements range from almost 50% to 94%. In one case ($t = 8$, $L = 0$ and $\alpha = 0.90$) an improvement of over 100% is achieved, so here underperformance changes to overperformance. All additional improvements turn out to be significant. Also, all total improvements, ranging from 63% to 96% (plus one case with 113%), are significant as well. Hence, using the correction function given in (4.7) results in a significantly better performance in this case study.

Figure 4.14 displays the aggregated attained service and relative deviation when using the P_2 criterion. Again, if we consider the 3153 demand streams individually, we find that some of them have an attained service that is higher than the desired one, while there are also demand streams with an attained service lower than the desired one. If we consider the aggregated service of all demand streams, we see that the desired service is not reached in the most cases. Also in case of the P_2 criterion using the corrections improves the attained service level. Using β' instead of β results in im-

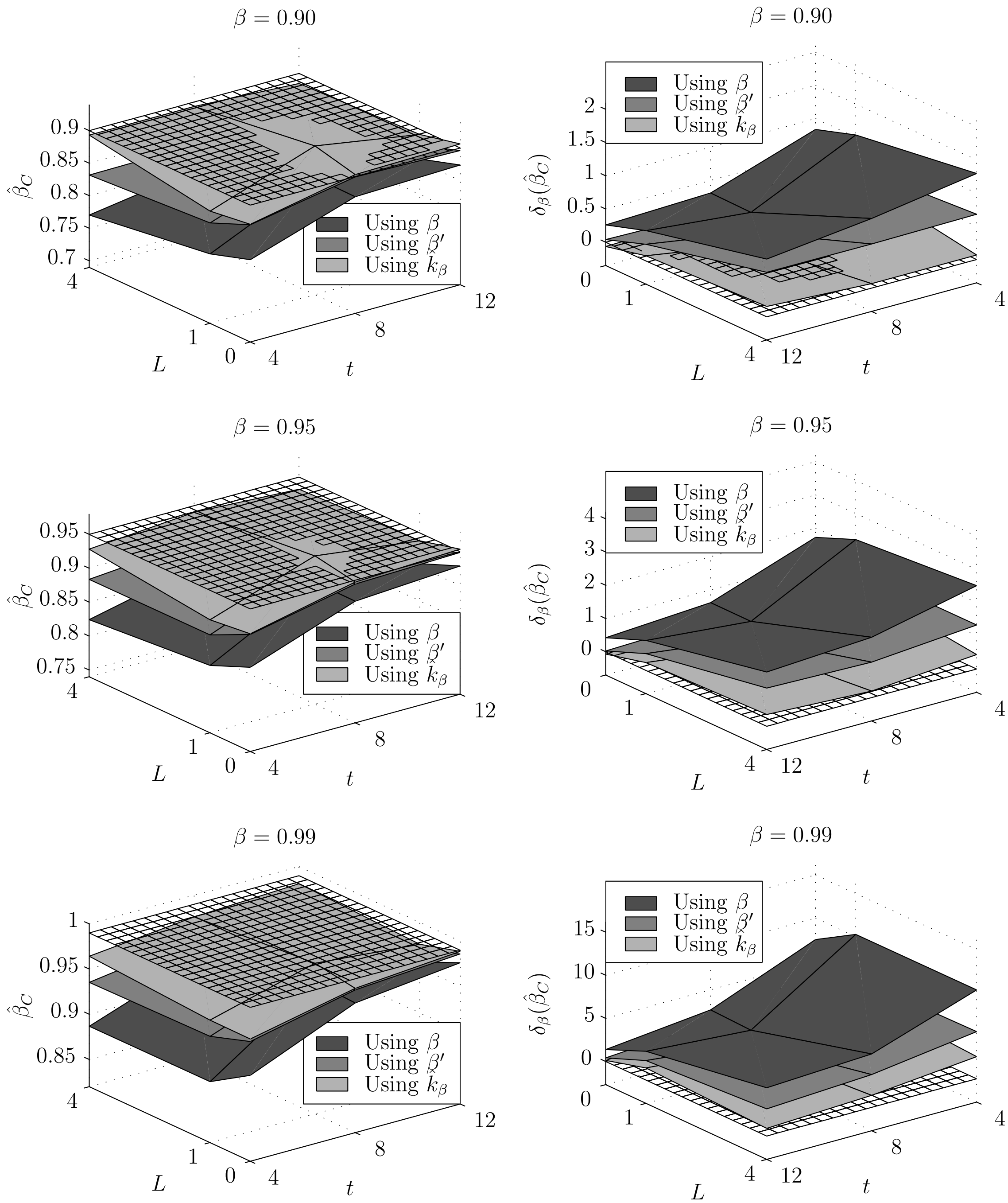


Figure 4.14: Attained service ($\hat{\beta}_C$) and relative deviation ($\delta_\beta(\hat{\beta}_C)$) using the P_2 criterion in the case study for.

provements ranging from almost 23% to 98%, plus one case ($t = 8$, $L = 1$, $\beta = 0.90$) in which the desired service level is attained; the attained service level is 0.9004.

Unfortunately, not all of these improvements are significant according to a t-test with a significance level of 5%; 10 out of the 27 cases are not: all the cases at which $t = 12$ plus the case at which $t = 8$, $L = 4$ and $\beta = 0.90$. If not only β' is used, but also the correction function given in (4.9), additional improvements, ranging from 2% up to 90% can be achieved in case the desired service level is not reached completely. In four more cases the desired service level is reached: $\alpha = 0.90, 0.95$, $t = 8, 12$ and $L = 0$. Only if $L = 4$ and $t = 4, t = 8$ these improvements turn out to be significant (for all the desired service levels). The total improvements range from 53% to 95% for the 22 cases in which the desired service level is not reached completely, and from 110% to 161% for the five cases in which the underperformance changes to overperformance. If we consider the total improvements all of them, except for $t = 12$, $L = 1$ and $\beta = 0.90, 0.99$, are significant. Hence we can state that using the adapted desired service β' and the correction function, depending on β, together improves the attained service level. Table 4.19 displays the extreme deviations.

Desired service level	Minimum attained service ($\delta_\beta(\hat{\beta}_C)$)	Maximum attained service ($\delta_\beta(\hat{\beta}_C)$)	Mean attained service ($\delta_\beta(\hat{\beta}_C)$)
$\beta = 0.90$			
Using β	0.7705 (1.295)	0.8739 (0.261)	0.8288 (0.712)
Using β'	0.8318 (0.682)	0.9004 (-0.004)	0.8648 (0.352)
Using $\hat{k}_\beta$	0.8624 (0.376)	0.9184 (-0.184)	0.8915 (0.085)
$\beta = 0.95$			
Using β	0.8244 (2.512)	0.9290 (0.420)	0.8808 (1.384)
Using β'	0.8836 (1.328)	0.9488 (0.024)	0.9147 (0.706)
Using $\hat{k}_\beta$	0.9064 (0.872)	0.9585 (-0.170)	0.9354 (0.292)
$\beta = 0.99$			
Using β	0.8868 (10.320)	0.9762 (1.380)	0.9348 (5.520)
Using β'	0.9353 (5.470)	0.9859 (0.410)	0.9608 (2.920)
Using $\hat{k}_\beta$	0.9551 (3.490)	0.9887 (0.130)	0.9736 (1.640)

Table 4.19: Extreme deviations from desired service level for $\beta \in \{0.90, 0.95, 0.99\}$ in the case study.

We also did the case study calculations with a rolling horizon and the results are similar; the significance results cannot be checked, since the outcomes on which the attained service levels are based are no longer independent.

4.5 Summary results

This chapter has considered the case of an (R, S) inventory control model with gamma demand and a service criterion. It is shown that using estimates in the determination of the order-up-to levels derived under the assumption that all parameters are known, leads to underperformance. If demand is exponentially distributed and the lead time is zero, it is shown that the desired service level is never reached. For the case of the P_1 criterion, Erlang demand, integer lead time and known shape parameter, we have derived closed-form expressions for the attained service and use these to show that the desired service is not met for higher values of α, i.e., $\alpha \geq 0.50$. For the most realistic situation treated in this chapter (demand is truly gamma distributed with unknown parameters) simulation is used to show that indeed underperformance exists for both the P_1 and P_2 criterion and the desired service level set at 0.90 or higher; these values are used in practice. Part of this underperformance could be solved by using α' instead of α for the P_1 criterion (improvements range from 18% to 80%) and β' instead of β for the P_2 criterion (improvements range from 16% to 99% in case the desired service level is not met; it is met in 17 out of 180 cases).

Further improvements are obtained by applying a multiplicative correction to the estimated order-up-to level. This correction is found using simulation and with help of nested regression a function is constructed. Using the correction function causes the attained service to reach the desired service even more closely. However, the desired service level is not reached completely, due to the fact that the correction functions are determined using the true value of ρ while in practice only $\hat{\rho}$ can be used. The additional (total) improvements range from 60% to 99% (76% to 99%) for the P_1 criterion. The results for the additional improvements of the P_2 criterion are a bit more complicated, since in some cases the desired service level is (almost) met without using the multiplicative correction. So in some cases the attained service level declines a little, but in general the additional improvements are between a few percent up to 94%. Total improvements range from 66% to 94%.

We also applied the corrections developed in this chapter to real demand data, that was provided by Involvation and the Dutch Ministry of Defence. In case of the P_1 criterion the total improvements range from 63% to 96% and in one case the desired service level is met. For the P_2 criterion the total improvements range from 53% to 95% in the 22 cases in which the desired service level is not met; it is met in the remaining 5 cases. This case study showed that, although our method has its limitations, it works well in practice.

Chapter 5

Mixed Erlang demand and random inventory control parameters

In this chapter the effect of randomness in the inventory control parameters under mixed Erlang demand is discussed. For some industries, like process industry, it is difficult to produce exactly the required replenishment quantity, due to, e.g., yield problems. Hence the optimal control parameter is more a target than an exact result. We provide a method to deal with this uncertainty in inventory control by considering the control parameters to be random, unlike classical inventory control. The method described in this chapter is for a periodic review, order-up-to policy (the (R, S) policy) under a service level constraint.

5.1 Introduction

Classical inventory theory assumes that demand can be stochastic, but that its distribution is known completely. Using that information it is possible to determine the control parameter settings of a chosen reordering system to either minimize the total costs or to reach a pre-specified service level. If the control parameters of the reordering system are chosen, they will not change, and, assuming that the demand is stationary, one can use these parameter settings, either to have minimal expected cost or to reach the desired service level in the long run. This chapter considers the effects that occur when these control parameters are random.

Chapters 3 and 4 treat the case that the demand parameters are random, which results in the order-up-to level, one of the control parameters, being random. This is not the only cause of randomness in the order-up-to level. Assume that indeed the demand distribution is completely known, so that the correct order-up-to level can be determined. Due to certain circumstances the value of this order-up-to level might not be attained exactly. One can think of manufacturing products: a certain number of products fail quality tests and have to be repaired or even discarded. The number of products that will be discarded is not exactly known beforehand, so there is some uncertainty in the amount of products that will be available in order to satisfy demand. This is often referred to as random yield; see Yano and Lee (1995) for a review of literature of lot sizing under random yield. Another example, from process industry, is the production of bacteria. It is possible to accurately aim the production at a certain level, but occasionally a batch is contaminated and the whole batch has to be destroyed.

The randomness of the length of the lead time is discussed in literature (see, e.g., Kaplan, 1970, Chapter 7 of Zipkin, 2000, or Silver and Robb, 2008); uncertainty in delivery times may be caused by, e.g., congestion, stock out at the supplier, or weather conditions in case of sea transport.

However, the fact that the length of the review period may be random is less present in literature. The source of this randomness might not be as intuitive compared to the lead time: one simply states that the inventory position is reviewed at a certain time interval, e.g., once a week, and then it is fixed. However, many chemical companies do face randomness in the length of the review period. These companies often have a cyclic production schedule: they produce products $1, \ldots, n$ and then start again with 1; see, e.g., Ashayeri et al. (2006). These cycles are also called campaigns. In order to determine the optimal inventory level such that a certain desired service level is attained, mostly an (R, S) or (R, s, S) policy is used; the latter implies that the inventory position is reviewed every R units of time and is replenished up to S if the inventory position is below s. The review of the inventory position typically takes place at the start of the campaign. The length of the review period may be exposed to variation, due to possible variation in the setup times or in the production times for each of the products in the campaign.

This chapter does not focus on the cause of the uncertainty but considers the consequences instead. In Chapters 3 and 4 the cause of the uncertainty of the order-up-to level is used to find a first improvement of the attained service. In this chapter, we do not only consider more control parameters being random (R and L are fixed in earlier chapters), but we also include more causes for the randomness of S, so this chapter is more general compared to Chapters 3 and 4.

For reasons of flexibility in estimating demand and mathematical convenience, we consider demand that follows a mixed Erlang distribution. Furthermore, we assume that an (R, S) policy is used, subject to a service level constraint; P_2 is considered. The demand distribution and inventory policy are discussed in Section 5.2. Section 5.3 shows that if the length of the review period R and/or the lead time L and/or the order-up-to level S are stochastic, the desired service level is not reached. Furthermore, an easy algorithm to solve this problem is described and, under certain conditions, the optimal order-up-to levels in case of random R, L, and S are found. Section 5.4 concludes this chapter.

5.2 Demand distribution and inventory policy

This chapter considers four random elements in determining the correct (target) order-up-to level:

1. Demand per period, with expected value $\mathbb{E}[D]$ and coefficient of variation $\mathbb{C}_D$;
2. Length of review period, with expected value $\mathbb{E}[R] = 1$ and coefficient of variation $\mathbb{C}_R$;
3. Lead time, with expected value $\mathbb{E}[L]$ and coefficient of variation $\mathbb{C}_L$;
4. Order-up-to level, with expected value $\mathbb{E}[S]$ and coefficient of variation $\mathbb{C}_S$.

The first, stochastic demand, is treated extensively in literature (see, e.g., Silver et al., 1998). This randomness is taken into account in all our calculations without elaborating on it. The remainder of this chapter discusses the three random elements remaining explicitly and we refer to those as the three sources of randomness. In the majority of the literature it is assumed that the length of the review period (in case of a periodic review policy) and the length of the lead time are known and fixed. Furthermore, it is assumed that the ordered amount will be received completely.

5.2.1 Inventory policy

We assume that an (R, S) inventory control policy with lead time L is used: each R units of time an order is placed such that the inventory position is equal to the order-up-to level S and that order is delivered L units of time later. This order-up-to level is chosen such that the fraction of demand that can be satisfied immediately, is at least

β. Demand that cannot be satisfied immediately, is backlogged. Mathematically, the attained service is defined as

$$\beta = 1 - \frac{\mathbb{E}\left[(D_{R+L} - S)^{+}\right] - \mathbb{E}\left[(D_L - S)^{+}\right]}{\mathbb{E}\left[D_R\right]}, \tag{5.1}$$

where D_{R+L} (D_L, D_R) is demand during the review period plus lead time (lead time, review period) and $(x)^{+} = \max(x, 0)$.

The possibility of R, L, and S being random implies that evaluating (5.1) could be more difficult. An easy 'solution' to the problem created by randomness is to assume that R and L are fixed and determine the order-up-to level using the expected value. However, if we ignore the extra variability caused by R and L, we may expect underperformance. Furthermore, also S may be random; if we ignore the randomness of S and substitute S by $\mathbb{E}[S]$, we may expect additional underperformance (see Chapters 3 and 4).

We still can use (5.1) to find the attained service level, but we need to take into account that not only the demand per period is random, but also R, L and S are.

5.2.2 Demand distribution

Equation (5.1) is elaborated on while assuming mixed Erlang distributed demand during the review period and during the lead time. If a variable X is mixed Erlang distributed, its pdf ($f_{\boldsymbol{\zeta}}(x)$) is

$$f_{\boldsymbol{\zeta}}(x) = p_1 \mu_1^{k_1} \frac{x^{k_1 - 1}}{(k_1 - 1)!} e^{-\mu_1 x} + p_2 \mu_2^{k_2} \frac{x^{k_2 - 1}}{(k_2 - 1)!} e^{-\mu_2 x},$$

where $\boldsymbol{\zeta}$ is a vector containing the six demand parameters p_1, p_2, k_1, k_2, μ_1, and μ_2. All parameters are nonnegative, $p_1 + p_2 = 1$, and $k_1, k_2 \in \mathbb{N}$. The demand during the review period has the demand parameters $\boldsymbol{\zeta} = [p_1, p_2, k_1, k_2, \mu_1, \mu_2]$; the demand during lead time has the demand parameters $\boldsymbol{\eta} = [q_1, q_2, l_1, l_2, \lambda_1, \lambda_1]$.

Also the order-up-to level S is assumed to be mixed Erlang distributed, with demand parameters $\boldsymbol{\xi} = [w_1, w_2, m_1, m_2, \rho_1, \rho_2]$.

Assuming that demand during review, demand during lead time, and the order-up-to level can be accurately fitted to a mixed Erlang distribution with demand parameters $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$, respectively, we can determine two of the three expected values in (5.1), namely $\mathbb{E}[D_R]$ and $\mathbb{E}[(D_L - S)^{+}]$:

$$\mathbb{E}\left[D_R\right] = p_1 \frac{k_1}{\mu_1} + p_2 \frac{k_2}{\mu_2}, \tag{5.2}$$

$$\begin{aligned}&\mathbb{E}\left[(D_L - S)^+\right]\\&\overset{S \text{ fixed}}{=} \sum_{j=1}^{2} q_j \left(\frac{l_j}{\lambda_j} \sum_{t=0}^{l_j} \frac{(\lambda_j S)^t}{t!} e^{-\lambda_j S} - S \sum_{t=0}^{l_j - 1} \frac{(\lambda_j S)^t}{t!} e^{-\lambda_j S} \right),\end{aligned} \tag{5.3}$$

$$\begin{aligned}&\mathbb{E}\left[(D_L - S)^+\right]\\&\overset{S \text{ random}}{=} \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\sum_{t=0}^{l_j} \frac{\rho_i^{m_i} l_j \lambda_j^{t-1} (m_i + t - 1)!}{(m_i - 1)! t! (\lambda_j + \rho_i)^{m_i + t}} \right.\\&\qquad\qquad\left. - \sum_{t=0}^{l_j - 1} \frac{\rho_i^{m_i} \lambda_j^t (m_i + t)!}{(m_i - 1)! t! (\lambda_j + \rho_i)^{m_i + t + 1}} \right).\end{aligned} \tag{5.4}$$

See Appendix E.2 for the derivation of (5.3) and (5.4).

We also need an expression for $\mathbb{E}\left[(D_{R+L} - S)^+\right]$. If D_{R+L} is mixed Erlang distributed, we would get similar expressions as in (5.3) and (5.4). However, the sum of two mixed Erlang distributed variables is, in general, not mixed Erlang distributed. The mixed Erlang distribution is a special case of the phase type distribution (see Appendix E.3) and the sum of two phase type distributed variables has again a phase type distribution. Hence, we know that D_{R+L} is phase type distributed with demand parameters $\boldsymbol{\tau}_{R+L}$ and T_{R+L}. If S is not assumed to be random, the expected amount of backlog at the end of the replenishment cycle is

$$\mathbb{E}\left[(D_{R+L} - S)^+\right] = \int_S^\infty x f_{T_{R+L}, \boldsymbol{\tau}_{R+L}}(x) dx - S\big(1 - F_{T_{R+L}, \boldsymbol{\tau}_{R+L}}(S)\big),$$

where $f_{T_{R+L}, \boldsymbol{\tau}_{R+L}}$ denotes the pdf of a phase type distribution. If S is random, the expected amount of backlog at the end of the replenishment cycle is

$$\begin{aligned}\mathbb{E}\left[(D_{R+L} - S)^+\right] = &\int_0^\infty \Big(\int_s^\infty x f_{T_{R+L}, \boldsymbol{\tau}_{R+L}}(x) dx \Big) f_{\boldsymbol{\xi}}(s) ds - \left(w_1 \frac{m_1}{\rho_1} + w_2 \frac{m_2}{\rho_2} \right)\\&+ \int_0^\infty s F_{T_{R+L}, \boldsymbol{\tau}_{R+L}}(s) f_{\boldsymbol{\xi}}(s) ds,\end{aligned}$$

with $f_{\boldsymbol{\xi}}$ the pdf of a mixed Erlang distribution with parameters $\boldsymbol{\xi}$. The derivation of these two expressions is in Appendix E.4, as are the algorithms needed to calculate the pdf and cdf of the phase type distribution. Note that these expressions still have integrals, which need to be numerically determined. This causes calculations to be slow.

5.2.3 Determining the order-up-to level

We find the distribution parameters in $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$ by assuming that we know the first two moments of the demand per period ($\mathbb{E}[D]$ and $\mathbb{C}_D$), the length of the review period ($\mathbb{E}[R]$ and $\mathbb{C}_R$) and the lead time ($\mathbb{E}[L]$ and $\mathbb{C}_L$). With help of those we can determine the first two moments of the demand during review, and demand during lead time:

$$\begin{aligned} \mathbb{E}[D_\ell] &= \mathbb{E}[\ell] \cdot \mathbb{E}[D] && \ell \in \{R, L\} \\ \mathbb{C}^2_{D_\ell} &= \mathbb{C}^2_D + \mathbb{C}^2_\ell + \mathbb{C}^2_D \mathbb{C}^2_\ell && \ell \in \{R, L\}. \end{aligned}$$

Next, we can use (E.1) to find the parameters of the mixed Erlang distribution belonging to D_R and D_L. Further, we assume that the order-up-to level S has expected value $\mathbb{E}[S]$ and coefficient of variation $\mathbb{C}_S$. If we fit a mixed Erlang distribution using these first two moments using (E.1), the order-up-to level S has parameters $\boldsymbol{\xi}$.

The attained backlogged demand is

$$\mathbb{E}\left[(D_{R+L} - S)^+\right] - \mathbb{E}\left[(D_L - S)^+\right] \tag{5.5}$$

and the desired backlogged demand is

$$(1 - \beta)\mathbb{E}[D_R], \tag{5.6}$$

where β is the desired service level. Note that (5.5) is decreasing in S, since the higher the order-up-to level is, the lower the amount of backlogged demand will be.

If S is assumed to be deterministic, we can use binary search to find the value for S such that (5.5) and (5.6) are equal. However, if we take possible randomness of S into account, finding a value for S is not straightforward. In fact, we would like to have values for $\mathbb{E}[S]$ and $\mathbb{C}_S$; having these values we should aim at realizing these values in practice. However, it is not realistic that we can choose a value for $\mathbb{C}_S$, since this value is a function of the production system itself. Being able to vary the expected value does seem reasonable, since simply ordering a little more or a little less will change the expected value. Hence, we decided to keep the coefficient of variation fixed and only vary the expected value in the binary search. The resulted expected value is then a *target* value for the order-up-to level instead of *the* value.

5.3 Validations and results

We have determined eight order-up-to levels for all combinations of different values for the eight demand and control parameters listed in Table 5.1. The eight order-up-to levels take account of the randomness in various degrees. First, we assume no randomness and the order-up-to level is denoted by S_{rls}. The indices show the three possible sources of randomness (length of review period, lead time and order-up-to level, respectively). Lower case letters indicate no randomness assumed, while upper case letters denote randomness. Next, we assume randomness in one of the three sources; the three order-up-to levels are S_{rlS}, S_{rLs} and S_{Rls}. Then, two sources of randomness are assumed; the order-up-to levels are S_{rLS}, S_{RlS} and S_{RLs}. Finally, we assume randomness in all three sources and this order-up-to level is S_{RLS}.

Demand/control parameter	Values used
$\mathbb{E}[D]$	10, 20
$\mathbb{C}_D$	0.25, 0.50, 0.75
$\mathbb{E}[R]$	1
$\mathbb{C}_R$	0.25, 0.50, 0.75
$\mathbb{E}[L]$	0.5, 1, 2, 4
$\mathbb{C}_L$	0.25, 0.50, 0.75
$\mathbb{C}_S$	0, 0.25, 030, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60
β	0.90, 0.95

Table 5.1: Values used in simulation.

The eight order-up-to levels are determined for all 3,888 combinations of demand/control parameters listed in Table 5.1 as described in Section 5.2.3. The correctness of the computer program and formulae is checked using simulation. For each of the $3,888$ combinations we generated $n = 10,000$ times the demand during review from a mixed Erlang distribution with parameters $\boldsymbol{\zeta}$, denoted by

$$d_{iR}, \quad i = 1, \ldots, n,$$

and $10,000$ times the demand during lead time from a mixed Erlang distribution with parameters $\boldsymbol{\eta}$, denoted by

$$d_{iL}, \quad i = 1, \ldots, n.$$

The demand during review plus lead time is the sum of these two:

$$d_{iR+L} = d_{iR} + d_{iL}, \quad i = 1, \ldots, n.$$

Since S is assumed to be random for four of the order-up-to levels, also $10,000$ values for each of S_{rlS}, S_{rLS}, S_{RlS}, and S_{RLS} need to be generated from a mixed Erlang distribution with parameters ξ; denote these order-up-to levels by $s_{i,rlS}$, $s_{i,rLS}$, $s_{i,RlS}$, and $s_{i,RLS}$. Note that $s_{i,rls} = S_{rls}$, $s_{i,rLs} = S_{rLs}$, $s_{i,Rls} = S_{Rls}$, and $s_{i,RLs} = S_{RLs}$, since there is no randomness in these order-up-to levels.

The simulated attained services $\hat{\beta}_{rls}, \ldots, \hat{\beta}_{RLS}$ for all eight order-up-to levels for each combination are then determined using the $n = 10,000$ demand observations:

$$\hat{\beta}_{\mathcal{C}} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\Big((d_{iR} + d_{iL} - s_{i,\mathcal{C}})^+ - (d_{iL} - s_{i,\mathcal{C}})^+\Big)}{\frac{1}{n}\sum_{i=1}^{n} d_{iR}}, \quad \mathcal{C} \in \{rls, \ldots, RLS\}.$$

We found that the simulated attained service levels are indeed very close to the desired service levels; using a standard t-test with a significance level of 5% we find that 95.07% of the $3,888 \cdot 8 = 31,104$ simulated attained service levels did not differ significantly from the desired service level, so we can state that the (target) order-up-to levels determined in the way described in Section 5.2.3 do result in reaching the desired service level.

The remainder of this section considers the implications of adding the three sources of randomness (length of review period, lead time, size of order-up-to level). Beforehand, we expect that randomness in each of the variables will lead to a higher (targeted) order-up-to level for the desired service level to be reached, since additional randomness causes variability in timing and quantities. Thus, in general, more inventory is needed to be able to cope with this variability.

Figure 5.1 displays the eight (target) order-up-to levels for various values of $\mathbb{C}_S$ and the attained service when using $S_{rls}, S_{rlS}, \ldots, S_{RLS}$ while all randomness is present. First consider the left graph. The target order-up-to level is the expected value of the order-up-to level when S is random. Note that if the order-up-to level is not assumed to be random, $\mathbb{C}_S$ does not have an effect on the order-up-to level, which is logical. These order-up-to levels are depicted with the solid markers. Further, comparing the two order-up-to levels for which only the randomness in S differs, e.g., S_{rLs} (solid squares) and S_{rLS} (open squares), one sees that the larger $\mathbb{C}_S$, the larger the order-up-to level, as we expected. Figure 5.1 also shows that if one takes the randomness of the length of the review period into account, while keeping the other sources of randomness fixed, the order-up-to level increases; compare the lines with the circles (R deterministic) to the lines with the diamonds (R random) and the lines

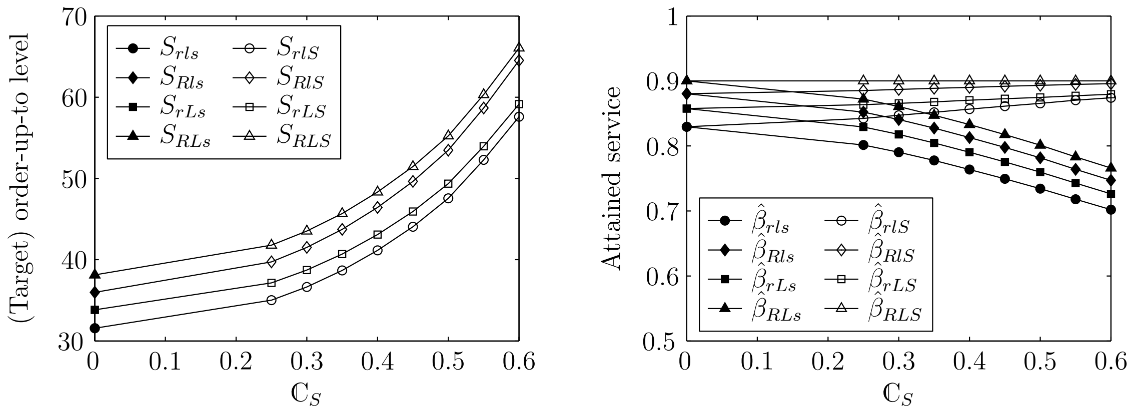


Figure 5.1: Eight order-up-to levels and the attained service for $\mathbb{E}[D] = 10$, $\mathbb{C}_D = 0.75$, $\mathbb{E}[R] = 1$, $\mathbb{C}_R = 0.5$, $\mathbb{E}[L] = 1$, $\mathbb{C}_L = 0.5$ and $\beta = 0.90$ for different values of $\mathbb{C}_S$.

with the squares (R deterministic) to the lines with the triangles (R random). The same holds for the randomness of the lead time; compare the lines with the circles (L deterministic) to the lines with the squares (L random) and the lines with the diamonds (L deterministic) to the lines with the triangles (L random).

The right graph in Figure 5.1 depicts the attained service levels, $\hat{\beta}_{rls}, \hat{\beta}_{rlS}, \ldots, \hat{\beta}_{RLS}$, when using the order-up-to levels $S_{rls}, S_{rlS}, \ldots, S_{RLS}$, while in fact the order-up-to level S_{RLS} should be used. So using that order-up-to level leads to reaching the desired service level (the line with the open triangles). If we ignore one, two, or all three sources of randomness, the desired service level is not reached. Note that the attained service is increasing in $\mathbb{C}_S$ when the randomness of S is taken into account, while it is decreasing in $\mathbb{C}_S$ when the randomness of S is not taken into account. The latter is logical, since the higher the coefficient of variation of S, the lower the attained service if we ignore that randomness. It is also trivial that the attained service is better when the randomness of S is taken into account (the lines with open markers) compared to the case that this randomness is ignored (the lines with solid markers). The increase of the attained service when $\mathbb{C}_S$ is increasing is not trivial. One might expect that when the randomness of S is taken into account, the attained service is independent of $\mathbb{C}_S$. However, Figure 5.1 clearly shows that the attained service increases with $\mathbb{C}_S$, except when using S_{RLS}; in that case we attain the desired service level for every value of $\mathbb{C}_S$. An intuitive explanation for this phenomenon might be that the larger $\mathbb{C}_S$ is, the larger the part of the underperformance is that can be explained by the variability of the order-up-to level. So, if this randomness is taken into account for low $\mathbb{C}_S$ the effect on the attained service level might be smaller than if it is taken into account for high $\mathbb{C}_S$.

These graphs can be made for all different combinations of $\mathbb{E}[D]$, $\mathbb{C}_D$, $\mathbb{E}[L]$, $\mathbb{C}_L$, $\mathbb{E}[R]$, $\mathbb{C}_R$, and β. They all have similar characteristics (the more variability, the higher the order-up-to level). However, the relative increase in the order-up-to level differs among the different combinations. Figures 5.2 and 5.3 provide two examples. Figure 5.2 displays a situation where the coefficient of variation of the length of the

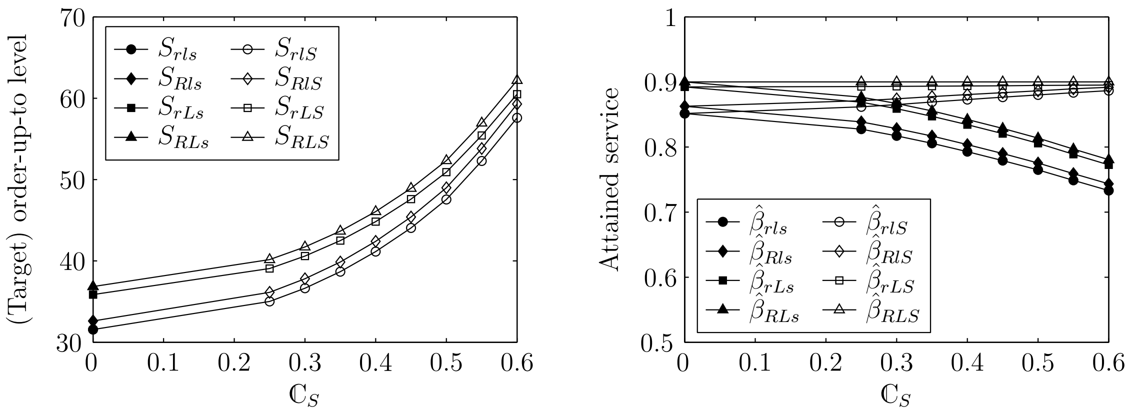


Figure 5.2: Eight order-up-to levels and the attained service for $\mathbb{E}[D] = 10$, $\mathbb{C}_D = 0.75$, $\mathbb{E}[R] = 1$, $\mathbb{C}_R = 0.25$, $\mathbb{E}[L] = 1$, $\mathbb{C}_L = 0.75$ and $\beta = 0.90$ for different values of $\mathbb{C}_S$.

review period is low compared to the coefficient of variation of the length of the lead time. Comparing the lines for which only the randomness of R differs (circles to diamonds and squares to triangles) one sees that those differences are smaller than the differences between the lines for which the randomness of L differs (circles to squares and diamonds to triangles). Figure 5.3 displays a situation where the coefficient of variation of the length of the review period is high compared to the coefficient of variation of the length of the lead time. Comparing the lines for which only the randomness of R differs (circles to diamonds and squares to triangles) one sees that those differences are larger than the differences between the lines for which the randomness of L differs (circles to squares and diamonds to triangles).

Considering Figures 5.1–5.3, one clearly sees that ignoring randomness while it is present leads to nonnegligible underperformance. In order to reach the desired service level the target order-up-to level needs to be considerably larger, up to more than twice as large. This implies that inventory level will increase considerably and that means more inventory costs. Whether increasing inventory level (and thus inventory costs) is justified by reaching the desired service level, is a subjective question, that every inventory manager needs to answer for himself. However, the method described

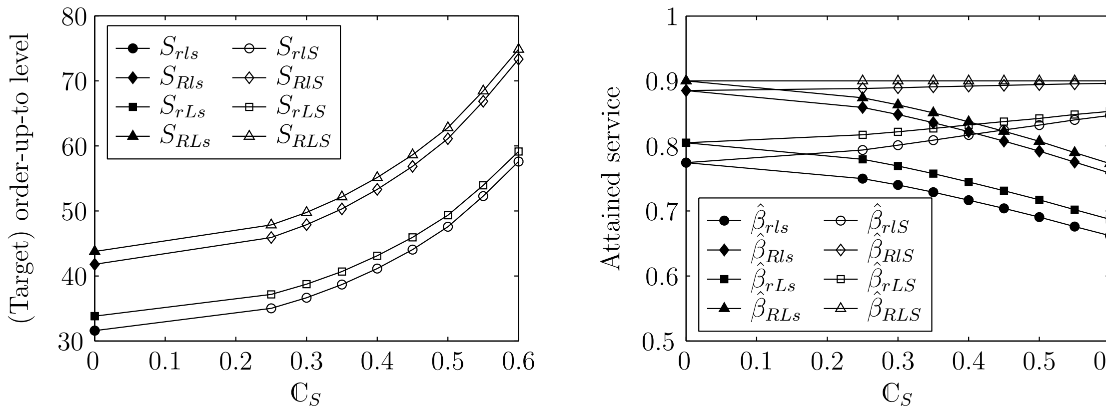


Figure 5.3: Eight order-up-to levels and the attained service for $\mathbb{E}[D] = 10$, $\mathbb{C}_D = 0.75$, $\mathbb{E}[R] = 1$, $\mathbb{C}_R = 0.75$, $\mathbb{E}[L] = 1$, $\mathbb{C}_L = 0.50$ and $\beta = 0.90$ different values of $\mathbb{C}_S$.

in this chapter provides a tool for that inventory manager to make this decision. Furthermore, it is often difficult to change the underlying reason of R, L and S being random. However, seeing how much can be gained in terms of inventory costs, the inventory manager might consider measures that seemed to be too expensive before. He might consider using the supplier that is more expensive, but has a less variable or even fixed lead time.

5.4 Summary results

This chapter has considered the effect of randomness in inventory control parameters. In most literature it is assumed that control parameters are fixed, while in fact they may be random. We have considered a periodic review, order-up-to level inventory control policy (the (R, S) policy), which leads to three control parameters: the length of the review period, the length of the lead time and the order-up-to level. A method to take these three sources of randomness into account is provided while assuming that the demand during the review period and demand during lead time is mixed Erlang distributed. The order-up-to levels for 3888 combinations of demand and control parameter values are determined, together with the attained service levels. Simulation is used to validate the correctness of the order-up-to levels.

It is shown that not taking the randomness of the inventory control parameters into account can lead to considerable underperformance. The figures in this chapter show attained service levels below 0.70, while the desired service level is 0.90: an underperformance of more than 20 percent points. The highest underperformance for all the 3888 combinations of demand and control parameters considered is 0.3659 (parameter setting $\mathbb{E}[D] = 10$, $\mathbb{C}_D = 0.25$, $\mathbb{E}[R] = 1$, $\mathbb{C}_R = 0.75$, $\mathbb{E}[L] = 1$, $\mathbb{C}_L = 0.25$, $\beta = 0.95$ and $\mathbb{C}_S = 0.60$).

Furthermore, the more randomness is taken into account and the more variable the control parameters are, the higher the order-up-to level needs to be. This confirms our a priori intuition and also is in line with the results of Chapters 3 and 4, in which we showed that the more variable the order-up-to level, the larger the underperformance is. We especially considered the randomness of the order-up-to level and showed with graphs that the order-up-to level rapidly increases when the coefficient of variation of the order-up-to level increases. This leads to higher inventory levels and, hence, higher inventory costs. One could argue that it might be wiser to investigate the source of the randomness thoroughly than to implement this method without a second thought.

A second limitation on implementing this method directly is that it assumes that the first two moments of demand, length of the review period, length of the lead time and the height of the order-up-to level are known with certainty. However, in real life, this will not be the case and we need to estimate them. This will probably again lead to underperformance (see Chapters 3 and 4).

The main idea of this method, taking account of the randomness of the inventory control parameters, has been applied to the inventory control in a big Dutch chemical company, although it is not exactly as described in this chapter. In that case the order-up-to level is assumed to follow a continuous uniform distribution, while demand during review, lead time and review plus lead time is mixed Erlang distributed. Using this approach has led to improving the inventory management of this company considerably.

Chapter 6

Conclusions and future research

This chapter discusses overall results and conclusions. Also ideas for future research are provided.

6.1 Conclusions

In the first part of this dissertation we construct a more flexible way to model demand using two modified shifted gamma distributions. The first one, the modified shifted gamma distribution with a point mass at zero, is constructed from a shifted gamma distribution by setting all negative realizations to zero. This leads to the point mass at zero, with probability of having a negative realization in the shifted gamma distribution we start from. The second one, the truncated shifted gamma distribution, is constructed from a shifted gamma distribution by ignoring all negative realizations. Using these distributions leads to nice expressions for the order-up-to levels that are easy to implement. Furthermore, the modified shifted gamma distribution with a point mass at zero can directly be used to model intermittent demand, since the possibility of having zero demand is not zero. Most commonly used demand distributions do not have this property and, therefore, other approaches are used to model intermittent demand. Furthermore, we prove that some results derived for both modified normal and modified gamma distributions hold in general. Finally, we show the attained service when using a regular gamma distribution or a shifted gamma distribution, while demand is actually modified shifted gamma distributed. We see that the attained service does not differ largely from the desired service level; in some cases it is a little below (maximal underperformance is 2.86 percent points) and in some cases it is a little above (maximal overperformance is 2.37 percent points). Still, not using the correct demand distribution does lead to not reaching the desired service level and a deviation of 2.5 percent points can be too large, depending on the market requirements.

The second part of this dissertation, Chapters 3–5, shows that not taking all existing randomness into account leads to serious underperformance. In Chapters 3 and 4 we focus on randomness in the order-up-to level that is caused by estimation of the demand parameters. These estimates are based on historical demand observations. The fewer the number of demand observations, the higher the variability in parameter estimates and hence the higher the variability in the order-up-to level, which is a function of the demand parameters in most inventory control policies. Intuitively this will lead to larger underperformance and it is shown that this indeed happens in case of normal and gamma distributed demand, if the desired service level is relatively large (at levels commonly used in practice). It is also shown that the larger the variability in demand, the larger the underperformance, and that the larger the desired service, the larger the relative underperformance. In Chapter 4 also lead time is included and the larger the lead time, the larger the underperformance. None of these results are counterintuitive.

Not only the existence and size of the underperformance is determined, also a method to find an order-up-to level that leads to (almost) attaining the desired service level is constructed. Both for the normal and gamma demand we analytically derive a first improvement under strong assumptions (only mean demand unknown in case of normal demand and exponential demand with zero lead time in case of gamma demand). Next to that, we use a technique we call nested regression to construct a correction function. In case of normal demand we find an additive correction function, depending on the coefficient of variation, the number of historical demand observations and the desired service level. If the true coefficient of variation is used, the desired service level is indeed reached. However, in real life the coefficient of variation needs to be estimated. Using this estimate leads to attaining the desired service level with a deviation of at most 0.5 percent points. In case of gamma demand we construct a multiplicative correction function depending on the coefficient of variation, the number of historical demand observations, the desired service level and the lead time. Using this function with the true value of the coefficient of variation leads to reaching the desired service level. If we use an estimate for the coefficient of variation, the desired service level is not reached in all cases and especially in case of a large coefficient of variation, the underperformance can still be substantial. However, using the correction function together with the first improvement does improve the attained service level significantly. The method of modeling demand according to a gamma distribution and using the first improvement and the correction function is applied to a case study. Not using any correction results in substantial underperformance, the first improvement increases the attained service, and using the correction function as well leads to almost reaching the desired service level. Hence, the case study shows that the method developed in Chapter 4 can also be used in practice.

The method for finding a correction to the order-up-to level as developed in Chapters 3 and 4 can be used more general. If we have one output variable that depends on multiple input variables and we use a full factorial design of the different values of those input variables, nested linear regression (see Appendix B) can be used to determine a function of the input variables that estimates the output variable. All three correction functions in this dissertation (determined using nested linear regression) have high coefficients of determination; they are all more than 0.99. Hence, we can state that the determined correction functions are very accurate in estimating the correction needed.

Chapter 5 does not only consider randomness in the order-up-to level, but also randomness in the other two inventory control parameters: length of the review period and length of the lead time. Furthermore, we do not reflect on the cause of the randomness in the three control parameters. This chapter studies the effect these three sources of randomness have on attaining the desired service level in case of mixed Erlang distributed demand during the review period and during the lead time and a mixed Erlang distributed order-up-to level. We have chosen the mixed Erlang distribution, since it is a flexible distribution that is mathematically tractable as well. We show analytically that ignoring the randomness leads, again, to substantial underperformance. Taking the randomness of the three control parameters into account leads to reaching the desired service level at the expense of high order-up-to levels, which result in high inventory costs. Hence, it could be wiser to try reducing the randomness than to mindlessly implement the new order-up-to levels. One could reduce the randomness in several ways, e.g., by choosing a more expensive but more reliable supplier to decrease the randomness in the lead time, or investing in new production equipment, which leads to less rejected products, hence less randomness in the order-up-to level. Also investing in a better forecast method could help, since also forecasting (estimation of the demand parameters) leads to randomness in the order-up-to level, as discussed in Chapters 3 and 4.

6.2 Ideas for future research

In this dissertation we consider the (R, S) inventory control policy, because of the relative easy derivation of the order-up-to levels. One main idea for future research is considering other inventory control policies. We expect to see similar results, especially regarding Part II: if randomness in the control parameters is not taking into account, the desired service level will not be reached and substantial underperformance arises.

Furthermore, in this dissertation we are interested in the attained service in the long run. In case of the cycle service criterion, it does not make a difference whether we consider the attained service over a finite horizon or whether we consider the long run attained service, but in case of the fill rate the expected attained service over a finite horizon can be shown to be higher than the long run attained service, see Thomas (2005). It might be interesting to study the attained service over a finite horizon in the case of unknown demand and control parameters, especially in the light of the assumption of a finite number of historical demand observations.

Chapter 2 considers two modified shifted gamma distributions to model demand in a setting with zero lead time. The work in this chapter is inspired by the work of Strijbosch and Moors (2006), who also assume zero lead time. It is already argued that zero lead time might seem very rigid, but that it could occur in practice, e.g., a supermarket could order after opening hours and that order is delivered the next morning before the supermarket opens again. However, the lead time will be positive in general and therefore one could expand the work done in Chapter 2 and in Strijbosch and Moors (2006) by including the possibility of a positive lead time.

One of the advantages of using the modified shifted gamma distribution with a point mass at zero is that it can be used to model intermittent demand. We have not elaborated on that in this dissertation; it is an idea for future research, including comparing using the modified shifted gamma distribution with a point mass at zero to using a compound distribution to model intermittent demand, e.g., the compound Bernoulli distribution, as described in Janssen et al. (1998).

Furthermore, in this chapter it is assumed that all demand parameters are known with certainty, while in practice this will not be the case. Hence, the demand parameters need to be estimated, for example by using the first three moments of the modified gamma distributions. In Chapters 3 and 4 we have seen that estimation of the demand parameters leads to underperformance, and, therefore, we expect that this will also occur in case of the modified gamma distributions. This is an interesting idea to elaborate on.

First shifting the gamma distribution and using this shifted gamma distribution to construct the modified shifted gamma distribution might be a little artificial, but the shifted gamma distribution can be used in the context of inventory management, albeit a little different compared to the use in this thesis. One could think of an inventory model in which the distribution of the forecast errors plays a role. One makes a forecast of the demand and the error belonging to that forecast can be either positive, when demand is higher than forecasted, or negative, when demand is lower than forecasted. The lowest demand possible is zero, so the forecast error can never be smaller than minus the forecast. In this situation the shifted gamma distribution arises naturally, since the forecast error can become infinitely large, but is bounded from below. The location parameter of the shifted gamma distribution (Δ) should be taken equal to the forecasted demand. So, if one would have an inventory model that needs the distribution of the forecast error as an input, using the shifted gamma distribution would be an interesting idea.

Part II considers the effect of randomness in the inventory control parameters, assuming that demand is stationary. In real life situations the demand pattern changes over time. A first study on the influence of forecasting on inventory control in case of non-stationary demand is performed by Strijbosch et al. (2010), which is accepted for publication. This is a very interesting topic to study further.

In Chapters 3 and 4 we consider sample statistics and the first two moments of the normal and gamma distribution to estimate demand during the review period and during the lead time. This is a very simple method of estimating demand (also known as (simple) moving average in forecasting), and is used in this dissertation because of the tractable results. There are more sophisticated methods to estimate or forecast demand. One widely used method is exponential smoothing; in this method the more recent demand observations are more important than the older demand observations and therefore it will react faster to changing demand patterns. Also exponential smoothing leads to tractable results and therefore most of the analytical derivations in this dissertation could also be performed with parameters that are estimated using exponential smoothing. Furthermore, we could also use more complex forecasting methods, e.g., ARIMA. If the method is calibrated in a good way, this will probably result in less underperformance. However, it is not likely that we will find any analytical results when using complex forecasting methods. Finally, we could also use maximum likelihood estimation to estimate the parameters. Unfortunately, this could lead to intractable results, depending on the distribution we want to fit on the historical demand observations. We could, of course, use simulation to find whether the maximum likelihood estimators lead to less underperformance.

In Chapters 3 and 4 correction functions are derived that depend, amongst others, on the coefficient of variation of demand. This is in both correction functions the only input variable that we do not know. Therefore, we estimate the coefficient of variation using the estimates for the mean and standard deviation. This estimate is not unbiased (the estimate of the standard deviation is not unbiased; see also Section 3.3.2), and therefore it is hard to believe that this is the best estimate for the coefficient of variation. Finding a better one could help in improving the attained service even further.

The correction functions in Chapters 3 and 4 are constructed using a technique we call nested linear regression. Of course, there are other methods known to estimate function prescriptions based on a set of values for input and output variables. Comparing these other methods to nested linear regression is an interesting idea to follow up.

In Chapter 3 the lead time is assumed to be zero. Zero lead time might not be unrealistic, but in most cases lead times will be positive. Hence, considering the possibility of non-zero (deterministic) lead time is a logical addition. We expect results analogous to Chapter 4, where non-zero lead time is considered. So, intuitively we expect to find that the larger the lead time, the lower the attained service and that the correction function needs to depend on the lead time, next to the coefficient of variation, the number of historical demand observations and the desired service level.

We could also include a random lead time in the situations described in Chapters 3 and 4. This will probably lead to even lower attained service levels if this randomness is ignored, as seen in Chapter 5. If it is taken into account, it will probably result in another input variable in the correction function, the coefficient of variation in the lead time, which needs to be estimated. It will complicate the analysis of these situations (normally distributed demand and gamma distributed demand) even further and probably no tractable results will exist, unless strict assumptions are made. However, since random lead times often appear in real life and the normal distribution is widely used (also the gamma distribution is used in practice), it will be a useful extension to the work as presented in Chapters 3 and 4.

The derivations in Chapter 5 are quite theoretical: we assume that we do know the means and coefficients of variation of demand, the length of the review period, the length of the lead time and the order-up-to level in order to show that randomness in the three control parameters leads to underperformance if ignored. In practice, we do not know these means and coefficients of variation with certainty, and we would need to estimate them using historical observations, probably leading again to even more underperformance. This model becomes very complex if we would like to include all unknown demand and control parameters and, therefore, tractable results cannot be expected. However, using simulation we could find the influences of all random elements in the model. Furthermore, if we make the ideas discussed in Chapter 5 more applicable in practice, we can do a comparison study of the three methods discussed in Chapters 3–5 and find which distribution, i.e., normal, gamma or mixed Erlang, is best to assume in different situations.

The three methods discussed in Chapters 3–5 all assume a continuous demand distribution, which implies that demand occurs in every review period (the probability of having zero demand is zero in a continuous distribution). However, a lot of types of products will have an intermittent demand pattern. If the possibility of zero demand is small or even negligible, then using a continuous demand distribution will not lead to large under- or overperformance. If this probability is not small, then we should use another method to estimate demand. One possibility is using the truncated normal distribution or truncated shifted gamma distribution, as discussed in Chapter 2. Another possibility is using a compound distribution. In both methods we need to estimate demand parameters and this estimation leads to variability in the order-up-to level. Since intermittent demand is essentially different from continuous demand, this does not necessarily lead to large underperformance and, hence, studying this situation might lead to interesting results and insights.

This dissertation is mainly about the effects of randomness in the demand and control parameters on the attained service. We have seen that these effects can be quite large and therefore the randomness cannot be ignored. Methods are provided to deal with the uncertainty in demand parameters, i.e., the correction functions constructed in Chapters 3 and 4. A really distribution-free approach can be found using robust optimization. In practice, one does not know the demand distribution and therefore assumes a distribution, e.g., the normal or gamma distribution. Not using the true distribution could lead to not reaching the desired service level (see Chapter 2), but whether under- or overperformance will occur, is not known beforehand;

note that both under- and overperformance are undesirable. If a distribution-free approach is used, one does not have the problem of choosing a good distribution. Applying robust optimization to inventory control is a relative new field of research. The main idea is using historical demand observations and a goodness-of-fit test to determine robust control parameters that can deal with uncertainty in demand and in the demand distribution. This might lead to an elegant solution to the problems that arise from the randomness in the demand and control parameters.

Appendix A

Derivations Chapter 2

A.1 Derivation of equations (2.1) and (2.2)

Let $f_{\rho,\theta}(x)$ and $F_{\rho,\theta}(x)$ denote the pdf and cdf of the regular gamma distribution and $f_{\rho,\theta,\Delta}(x)$ and $F_{\rho,\theta,\Delta}(x)$ the pdf and cdf of the shifted gamma distribution. We know that

$$\int_x^\infty (z-x) f_{\rho,\theta}(z) dz = \rho\theta[1-F_{\rho+1,\theta}(x)] - x[1-F_{\rho,\theta}(x)],$$
$$x f_{\rho,\theta}(x) = \theta\rho f_{\rho+1,\theta}(x).$$

We can rewrite these equations to the shifted gamma distribution using that $f_{\rho,\theta,\Delta}(x) = f_{\rho,\theta}(x+\Delta)$ and $F_{\rho,\theta,\Delta}(x) = F_{\rho,\theta}(x+\Delta)$ (or $f_{\rho,\theta,\Delta}(x-\Delta) = f_{\rho,\theta}(x)$ and $F_{\rho,\theta,\Delta}(x-\Delta) = F_{\rho,\theta}(x)$):

$$\int_w^\infty (y-w) f_{\rho,\theta}(y) dy = \rho\theta[1-F_{\rho+1,\theta}(w)] - w[1-F_{\rho,\theta}(w)],$$
$$\int_w^\infty (y-w) f_{\rho,\theta,\Delta}(y-\Delta) dy = \rho\theta[1-F_{\rho+1,\theta,\Delta}(w-\Delta)] - w[1-F_{\rho,\theta,\Delta}(w-\Delta)],$$
$$\text{Substitute } z = y-\Delta, y = z+\Delta, dy = dz \text{ (since } \tfrac{dy}{dz} = 1),$$
$$\int_{w-\Delta}^\infty (z+\Delta-w) f_{\rho,\theta,\Delta}(z) dz = \rho\theta[1-F_{\rho+1,\theta,\Delta}(w-\Delta)] - w[1-F_{\rho,\theta,\Delta}(w-\Delta)],$$
$$\text{Substitute } x = w-\Delta, w = x+\Delta,$$
$$\int_x^\infty (z-x) f_{\rho,\theta,\Delta}(z) dz = \rho\theta[1-F_{\rho+1,\theta,\Delta}(x)] - (x+\Delta)[1-F_{\rho,\theta,\Delta}(x)],$$

$$
\begin{aligned}
(x+\Delta) f_{\rho,\theta}(x+\Delta) &= \theta\rho f_{\rho+1,\theta}(x+\Delta), \\
(x+\Delta) f_{\rho,\theta,\Delta}(x) &= \theta\rho f_{\rho+1,\theta,\Delta}(x), \\
x f_{\rho,\theta,\Delta}(x) &= \theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta f_{\rho,\theta,\Delta}(x). \qquad \text{(A.1)}
\end{aligned}
$$

In one of the derivations in Chapter 2 we need to rewrite $x^2 f_{\rho,\theta,\Delta}(x)$; we can obtain that result by first multiplying both the left-hand side and right-hand side of (A.1) by x and then applying (A.1) to the right-hand side:

$$
\begin{aligned}
x \cdot (x f_{\rho,\theta,\Delta}(x)) &= x \cdot \big(\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta f_{\rho,\theta,\Delta}(x)\big) \qquad \text{(A.2)} \\
&= \theta\rho x f_{\rho+1,\theta,\Delta}(x) - \Delta x f_{\rho,\theta,\Delta}(x), \\
x^2 f_{\rho,\theta,\Delta}(x) &= \theta\rho[\theta(\rho+1) f_{\rho+2,\theta,\Delta}(x) - \Delta f_{\rho+1,\theta,\Delta}(x)] \\
&\quad - \Delta[\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta f_{\rho,\theta,\Delta}(x)], \\
x^2 f_{\rho,\theta,\Delta}(x) &= \theta^2\rho(\rho+1) f_{\rho+2,\theta,\Delta}(x) - 2\Delta\theta\rho f_{\rho+1,\theta,\Delta}(x) + \Delta^2 f_{\rho,\theta,\Delta}(x). \qquad \text{(A.3)}
\end{aligned}
$$

In another derivation in Chapter 2 we need to rewrite $x^3 f_{\rho,\theta,\Delta}(x)$; we can obtain that result by first multiplying both the left-hand side and right-hand side of (A.3) by x and then applying (A.1) to the right-hand side:

$$
\begin{aligned}
x \cdot (x^2 f_{\rho,\theta,\Delta}(x)) &= x \cdot \big(\theta^2\rho(\rho+1) f_{\rho+2,\theta,\Delta}(x) - 2\Delta\theta\rho f_{\rho+1,\theta,\Delta}(x) + \Delta^2 f_{\rho,\theta,\Delta}(x)\big), \\
x^3 f_{\rho,\theta,\Delta}(x) &= \theta^2\rho(\rho+1) x f_{\rho+2,\theta,\Delta}(x) - 2\Delta\theta\rho x f_{\rho+1,\theta,\Delta}(x) + \Delta^2 x f_{\rho,\theta,\Delta}(x) \\
&= \theta^2\rho(\rho+1)\big(\theta(\rho+2) f_{\rho+3,\theta,\Delta}(x) - \Delta f_{\rho+2,\theta,\Delta}(x)\big) \\
&\quad - 2\Delta\theta\rho\big(\theta(\rho+1) f_{\rho+2,\theta,\Delta}(x) - \Delta f_{\rho+1,\theta,\Delta}(x)\big) \\
&\quad + \Delta^2\big(\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta f_{\rho,\theta,\Delta}(x)\big) \\
&= \theta^3\rho(\rho+1)(\rho+2) f_{\rho+3,\theta,\Delta}(x) - 3\Delta\theta^2\rho(\rho+1) f_{\rho+2,\theta,\Delta}(x) \\
&\quad + 3\Delta^2\theta\rho f_{\rho+1,\theta,\Delta}(x) - \Delta^3 f_{\rho,\theta,\Delta}(x).
\end{aligned}
$$

A.2 First three moments of shifted gamma distribution

In this section we derive the first three moments of the shifted gamma distribution and use these to write the three demand parameters as a function of the first three moments. If X is distributed according to a regular gamma distribution with parameters ρ and θ (the shifted gamma distribution with ρ, θ, and $\Delta = 0$), the kth moment is given by (see, e.g., Evans et al., 2000):

$$\mathbb{E}\left[X^k\right] = \theta^k \frac{\Gamma(\rho + k)}{\Gamma(\rho)},$$

where $\Gamma(x)$ is the gamma function. Note that

$$\Gamma(x+1) = x\Gamma(x),$$

hence the first three moments of a regular gamma distributed variable are

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \theta^1 \frac{\Gamma(\rho + 1}{\Gamma(\rho)} = \theta \frac{\rho\Gamma(\rho)}{\Gamma(\rho)} = \theta\rho, \\
\mathbb{E}\left[X^2\right] &= \theta^2 \frac{\Gamma(\rho + 2)}{\Gamma(\rho)} = \theta^2 \frac{(\rho+1)\Gamma(\rho+1)}{\Gamma(\rho)} = \theta^2 \frac{(\rho+1)\rho\Gamma(\rho)}{\Gamma(\rho)} = \theta^2(\rho+1)\rho, \\
\mathbb{E}\left[X^3\right] &= \theta^3 \frac{\Gamma(\rho + 3)}{\Gamma(\rho)} = \theta^3 \frac{(\rho+2)(\rho+1)\rho\Gamma(\rho)}{\Gamma(\rho)} = \theta^3(\rho+2)(\rho+1)\rho.
\end{aligned}$$

Now let us consider a random variable X that is distributed according to a shifted gamma distribution with parameters ρ, θ, and Δ. The first moment of X is

$$\begin{aligned}
\mathbb{E}\left[X\right] &= \int_{-\infty}^{\infty} x f_{\rho,\theta,\Delta}(x) dx = \int_{-\Delta}^{\infty} x f_{\rho,\theta,\Delta}(x) dx = \int_{-\Delta}^{\infty} x f_{\rho,\theta}(x+\Delta) dx \\
&\overset{y=x+\Delta}{=} \int_{0}^{\infty} (y-\Delta) f_{\rho,\theta}(y) dy = \int_{0}^{\infty} y f_{\rho,\theta}(y) dy - \Delta \int_{0}^{\infty} f_{\rho,\theta}(y) dy \\
&= \theta\rho - \Delta.
\end{aligned}$$

The second moment of X is

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= \int_{-\infty}^{\infty} x^2 f_{\rho,\theta,\Delta}(x) dx = \int_{-\Delta}^{\infty} x^2 f_{\rho,\theta,\Delta}(x) dx = \int_{-\Delta}^{\infty} x^2 f_{\rho,\theta}(x+\Delta) dx \\
&\overset{y=x+\Delta}{=} \int_{0}^{\infty} (y-\Delta)^2 f_{\rho,\theta}(y) dy = \int_{0}^{\infty} (y^2 - 2y\Delta + \Delta^2) f_{\rho,\theta}(y) dy \\
&= \int_{0}^{\infty} y^2 f_{\rho,\theta}(y) dy - 2\Delta \int_{0}^{\infty} y f_{\rho,\theta}(y) dy + \Delta^2 \int_{0}^{\infty} f_{\rho,\theta}(y) dy \\
&= \theta^2(\rho+1)\rho - 2\Delta\theta\rho + \Delta^2.
\end{aligned}$$

The third moment of X is

$$\begin{aligned}
\mathbb{E}\left[X^3\right] &= \int_{-\infty}^{\infty} x^3 f_{\rho,\theta,\Delta}(x)dx = \int_{-\Delta}^{\infty} x^3 f_{\rho,\theta,\Delta}(x)dx = \int_{-\Delta}^{\infty} x^3 f_{\rho,\theta}(x+\Delta)dx \\
&\overset{y=x+\Delta}{=} \int_0^{\infty} (y-\Delta)^3 f_{\rho,\theta}(y)dy = \int_0^{\infty} (y^3 - 3y^2\Delta + 3y\Delta^2 - \Delta^3) f_{\rho,\theta}(y)dy \\
&= \int_0^{\infty} y^3 f_{\rho,\theta}(y)dy - 3\Delta \int_0^{\infty} y^2 f_{\rho,\theta}(y)dy + 3\Delta^2 \int_0^{\infty} y f_{\rho,\theta}(y)dy \\
&\quad - \Delta^3 \int_0^{\infty} y f_{\rho,\theta}(y)dy \\
&= \theta^3(\rho+2)(\rho+1)\rho - 3\Delta\theta^2(\rho+1)\rho + 3\Delta^2\theta\rho - \Delta^3.
\end{aligned}$$

Let us denote the first, second and third moment by μ, μ_2 and μ_3, so we have:

$$\begin{aligned}
\mu &= \theta\rho - \Delta, \\
\mu_2 &= \theta^2\rho^2 + \theta^2\rho - 2\Delta\theta\rho + \Delta^2, \\
\mu_3 &= \theta^3\rho^3 + 3\theta^3\rho^2 + 2\theta^3\rho - 3\Delta\theta^2\rho^2 - 3\Delta\theta^2\rho + 3\Delta^2\theta\rho - \Delta^3.
\end{aligned}$$

We now have these three moments as functions of the demand parameters ρ, θ and Δ; it could be useful to have the demand parameters as functions of μ, μ_2 and μ_3 (see 2.5). With a little rewriting we can obtain such functions. First, let us consider μ:

$$\mu = \rho\theta - \Delta \qquad \Rightarrow \qquad \theta\rho = \mu + \Delta \qquad \Rightarrow \qquad \rho = \frac{\mu+\Delta}{\theta}.$$

Next, substitute all ρ in the expression for μ_2 by $\frac{\mu+\Delta}{\theta}$ (or, equivalently, all $\theta\rho$ by $\mu + \Delta$):

$$\begin{aligned}
\mu_2 &= (\mu+\Delta)^2 + \theta(\mu+\Delta) - 2\Delta(\mu+\Delta) + \Delta^2 \\
&= \mu^2 + 2\mu\Delta + \Delta^2 + \theta\mu + \theta\Delta - 2\mu\Delta - 2\Delta^2 + \Delta^2 \\
&= \theta\mu + \theta\Delta + \mu^2.
\end{aligned}$$

Analogously for the expression of μ_3:

$$\begin{aligned}
\mu_3 &= (\mu+\Delta)^3 + 3\theta(\mu+\Delta)^2 + 2\theta^2(\mu+\Delta) \\
&\qquad - 3\Delta(\mu+\Delta)^2 - 3\Delta\theta(\mu+\Delta) + 3\Delta^2(\mu+\Delta) - \Delta^3 \\
&= \mu^3 + 3\mu^2\Delta + 3\mu\Delta^2 + \Delta^3 + 3\theta\mu^2 + 6\theta\mu\Delta + 3\theta\Delta^2 + 2\theta^2\mu + 2\theta^2\Delta \\
&\qquad - 3\Delta\mu^2 - 6\mu\Delta^2 - 3\Delta^3 - 3\Delta\theta\mu - 3\theta\Delta^2 + 3\Delta^2\mu + 3\Delta^3 - \Delta^3 \\
&= \mu^3 + 3\theta\mu^2 + 2\theta^2\mu + 3\theta\mu\Delta + 2\theta^2\Delta.
\end{aligned}$$

Now we can rewrite μ_2:

$$\mu_2 = \theta\mu + \theta\Delta + \mu^2 \qquad \Rightarrow \qquad \theta\Delta = \mu_2 - \mu^2 - \theta\mu$$
$$\Rightarrow \qquad \Delta = \frac{\mu_2 - \mu^2 - \theta\mu}{\theta} = \frac{\mu_2 - \mu^2}{\theta} - \mu.$$

Note that we can use this to rewrite the expression for ρ:

$$\rho = \frac{\mu + \frac{\mu_2 - \mu^2}{\theta} - \mu}{\theta} = \frac{\mu_2 - \mu^2}{\theta^2} = \frac{\sigma^2}{\theta^2}.$$

Also substitute $\Delta = \frac{\mu_2-\mu^2}{\theta} - \mu$ in the expression for μ_3:

$$\begin{aligned}\mu_3 &= \mu^3 + 3\theta\mu^2 + 2\theta^2\mu + 3\theta\mu\left(\tfrac{\mu_2-\mu^2}{\theta} - \mu\right) + 2\theta^2\left(\tfrac{\mu_2-\mu^2}{\theta} - \mu\right)\\ &= \mu^3 + 3\theta\mu^2 + 2\theta^2\mu + 3\mu\mu_2 - 3\mu^3 - 3\theta\mu^2 + 2\theta\mu_2 - 2\theta\mu^2 - 2\theta^2\mu\\ &= -2\mu^3 + 3\mu\mu_2 + 2\theta\mu_2 - 2\theta\mu^2 = -2\mu^3 + 3\mu\mu_2 + 2\theta(\mu_2 - \mu^2).\end{aligned}$$

Rewriting this last equation provides an expression of θ that solely depends on the first three moments of X:

$$2\theta(\mu_2 - \mu^2) = \mu_3 + 2\mu^3 - 3\mu\mu_2 \qquad \Rightarrow \qquad \theta = \frac{\mu_3 + 2\mu^3 - 3\mu\mu_2}{2(\mu_2 - \mu^2)}.$$

The expressions for Δ and ρ follow easily when θ is substituted with the expression above:

$$\Delta = (\mu_2 - \mu^2)\frac{2(\mu_2 - \mu^2)}{\mu_3 + 2\mu^3 - 3\mu\mu_2} = \frac{2(\mu_2 - \mu^2)^2}{\mu_3 + 2\mu^3 - 3\mu\mu_2},$$
$$\rho = (\mu_2 - \mu^2)\left(\frac{2(\mu_2 - \mu^2)}{\mu_3 + 2\mu^3 - 3\mu\mu_2}\right)^2 = \frac{4(\mu_2 - \mu^2)^3}{(\mu_3 + 2\mu^3 - 3\mu\mu_2)^2}.$$

Appendix B

Nested linear regression

Nested regression is used in Chapters 3 and 4 to estimate a correction function for the order-up-to level. To our knowledge, Strijbosch and Moors (1999) are the first to use this technique. The method is explained in this appendix.

We start with a dependent variable Y that depends on n independent variables $X_1, X_2, \ldots, X_n$. The different values of the independent variables (denoted by $\mathcal{X}_i$) form a full factorial design, so every combination of different values of $X_1, X_2, \ldots, X_n$ exists.

Example B.1 (Full factorial design)
The sets of possible values of X_1, X_2 and X_3 are $\mathcal{X}_1 = \{10, 20, 30\}$, $\mathcal{X}_2 = \{0.90, 0.95\}$ and $\mathcal{X}_3 = \{0.4, 0.8\}$. So there are $3 \cdot 2 \cdot 2 = 12$ combinations of X_1, X_2, X_3 for which a value of Y is known, i.e., the combinations:

X_1	X_2	X_3
10	0.90	0.4
10	0.90	0.8
10	0.95	0.4
10	0.95	0.8

X_1	X_2	X_3
20	0.90	0.4
20	0.90	0.8
20	0.95	0.4
20	0.95	0.8

X_1	X_2	X_3
30	0.90	0.4
30	0.90	0.8
30	0.95	0.4
30	0.95	0.8

□

The estimation process has n steps and in each of the steps one independent variable is chosen to be regressed on Y (in the first step) or on the estimated coefficients of the previous step. The regression is repeated for all combinations of the values of the variables that are not chosen that step or any of the previous steps. The estimated coefficients therefore depend on the non-chosen variables and this dependency is used in the following step(s). We follow this complicated procedure, since we want to have the possibility to choose a power to which the chosen independent variable is raised.

Below the steps are described in more detail. In this description we choose X_1 in the first step, X_2 in the second step, X_3 in the third, etcetera, but of course any order of choosing the independent variables is possible. Furthermore, let $\boldsymbol{X_i}$ denote the vector $[X_i, X_{i+1}, \ldots, X_n]'$ and $\boldsymbol{X}$ denote $[X_1, X_2, \ldots, X_n]'$.

Step 1 For every combination of values of $X_2, \ldots, X_n$ the value of Y is estimated depending on X_1. We assume that the relation has the following form:

$$Y(\boldsymbol{X}) = \gamma_0(r, \boldsymbol{X_2}) + \gamma_1(r, \boldsymbol{X_2})X_1^r + \varepsilon.$$

Note that the coefficients depend on the non-chosen variables and on the power to which X_1 is raised. If we would not have a constant (γ_0), we could use logarithms to find the value of r with help of linear regression, but this is not possible in the current form. Therefore, the coefficients are estimated for a set of values of r (denoted by $\mathcal{R}$), and the value for which the sum of squared errors (SSE for short) is minimized is chosen. For each value of r a regression equation $\hat{Y}(r, \boldsymbol{X}) = g_0(r, \boldsymbol{X_2}) + g_1(r, \boldsymbol{X_2})X_1^r$ is estimated. Then the best value of r, denoted by $\hat{r}$, is found by solving

$$\hat{r} = \operatorname{argmin}_{r\in\mathcal{R}} \sum_{X_1\in\mathcal{X}_1} \sum_{X_2\in\mathcal{X}_2} \cdots \sum_{X_n\in\mathcal{X}_n} \Big(Y(\boldsymbol{X}) - \hat{Y}(r, \boldsymbol{X})\Big)^2 .$$

Step 2 In this step the coefficients g_0 and g_1 are regressed on the independent variable X_2 for each of the combinations of $X_3, \ldots, X_n$. The relations are supposed to have the following form:

$$\begin{aligned} g_0 &= \gamma_{00}(r_0, \boldsymbol{X_3}) + \gamma_{01}(r_0, \boldsymbol{X_3})X_2^{r_0} + \varepsilon, \\ g_1 &= \gamma_{10}(r_1, \boldsymbol{X_3}) + \gamma_{11}(r_1, \boldsymbol{X_3})X_2^{r_1} + \varepsilon. \end{aligned}$$

For each value of r_0 and r_1 a regression equation $\hat{g}_i(r_i, \boldsymbol{X_2}) = g_{i0}(r_i, \boldsymbol{X_3}) + g_{i1}(r_i, \boldsymbol{X_3})X_2^{r_i}$ ($i \in \{0, 1\}$) is estimated and the best values for r_0 and r_1 are found by minimizing the SSE:

$$\begin{aligned} \hat{r}_0 &= \operatorname{argmin}_{r_0\in\mathcal{R}} \sum_{X_2\in\mathcal{X}_2} \cdots \sum_{X_n\in\mathcal{X}_n} \left(g_0(\boldsymbol{X_2}) - \hat{g}_0(r_0, \boldsymbol{X_2})\right)^2 , \\ \hat{r}_1 &= \operatorname{argmin}_{r_1\in\mathcal{R}} \sum_{X_2\in\mathcal{X}_2} \cdots \sum_{X_n\in\mathcal{X}_n} \left(g_1(\boldsymbol{X_2}) - \hat{g}_1(r_1, \boldsymbol{X_2})\right)^2 . \end{aligned}$$

Step 3 In this step the coefficients g_{ij} $(i, j \in \{0, 1\})$ are regressed on the independent variable X_3 for each of the combinations of $X_4, \ldots, X_n$. The relations are supposed to have the following form:

$$g_{ij} = \gamma_{ij0}(r_{ij}, \boldsymbol{X_4}) + \gamma_{ij1}(r_{ij}, \boldsymbol{X_4}) X_3^{r_{ij}} + \varepsilon.$$

For each value of r_{ij} a regression equation $\hat{g}_{ij}(r_{ij}, \boldsymbol{X_3}) = g_{ij0}(r_{ij}, \boldsymbol{X_4}) + g_{ij1}(r_{ij}, \boldsymbol{X_4}) X_3^{r_{ij}}$ is estimated and the best values for r_{ij} are found by minimizing the SSE:

$$\hat{r}_{ij} = \text{argmin}_{r_{ij} \in \mathcal{R}} \sum_{X_3 \in \mathcal{X}_3} \cdots \sum_{X_n \in \mathcal{X}_n} \left(g_{ij}(\boldsymbol{X_3}) - \hat{g}_{ij}(r_{ij}, \boldsymbol{X_3}) \right)^2 .$$

These steps are repeated in a similar way until the last step n.

Step $\boldsymbol{n}$ In the previous step coefficients $g_{i_1,\ldots,i_{n-1}}$ $(i_1, \ldots, i_{n-1} \in \{0, 1\})$ were estimated. In this last step we apply linear regression one more time. The relations are supposed to have the following form:

$$g_{i_1,\ldots,i_{n-1}} = \gamma_{i_1,\ldots,i_{n-1}0}(r_{i_1,\ldots,i_{n-1}}) + \gamma_{i_1,\ldots,i_{n-1}1}(r_{i_1,\ldots,i_{n-1}}) X_n^{r_{i_1,\ldots,i_{n-1}}} + \varepsilon.$$

For each value of $r_{i_1,\ldots,i_{n-1}}$ a regression equation is estimated, resulting in $\hat{g}_{i_1,\ldots,i_{n-1}} = g_{i_1,\ldots,i_{n-1}0}(r_{i_1,\ldots,i_{n-1}}) + g_{i_1,\ldots,i_{n-1}1}(r_{i_1,\ldots,i_{n-1}}) X_n^{r_{i_1,\ldots,i_{n-1}}}$ and the values for $r_{i_1,\ldots,i_{n-1}}$ are chosen such that the SSE is minimized:

$$\hat{r}_{i_1,\ldots,i_{n-1}} = \text{argmin}_{r_{i_1,\ldots,i_{n-1}} \in \mathcal{R}} \sum_{X_n \in \mathcal{X}_n} \left(g_{i_1,\ldots,i_{n-1}}(X_n) - \hat{g}_{i_1,\ldots,i_{n-1}}(r_{i_1,\ldots,i_{n-1}}, X_n) \right)^2 .$$

Finally, the estimates for the coefficients are substituted in each other and we obtain a formula which can be used to estimate Y, depending on $X_1, \ldots, X_n$:

$$\begin{aligned}
\hat{Y} &= g_0 + g_1 X_1^r \\
&= \left(g_{00} + g_{01} X_2^{r_0} \right) + \left(g_{10} + g_{11} X_2^{r_1} \right) X_1^r \\
&= \left(\left(g_{000} + g_{001} X_3^{r_{00}} \right) + \left(g_{010} + g_{011} X_3^{r_{01}} \right) X_2^{r_0} \right) \\
&\quad + \left(\left(g_{100} + g_{101} X_3^{r_{10}} \right) + \left(g_{110} + g_{111} X_3^{r_{11}} \right) X_2^{r_1} \right) X_1^r \\
&= \ldots
\end{aligned}$$

In order to clarify this procedure we will show how we estimated (3.15).

Example B.2 (Nested regression)
In Section 3.3.4 we try to find a good formula for the correction needed to attain the desired service level, denoted by k_s. This correction depends on ν, t and β (in the calculations we use $1-\beta$). We will first choose ν, then t and finally β.

Step 1 In the first step ν is regressed on k_s. The relation is assumed to have the following form:

$$k_s(\nu, t, \beta) = \gamma_0(r, t, \beta) + \gamma_1(r, t, \beta)\nu^r + \varepsilon. \tag{B.1}$$

Figure B.1 displays the relation between ν and k_s for three different combinations of t and β. It shows that the assumed relation seems to fit quite well to the actual

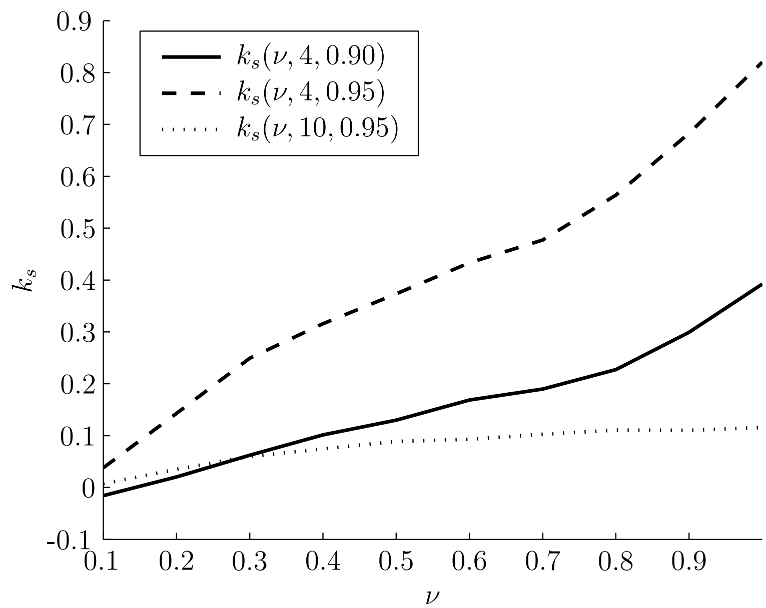


Figure B.1: $k_s\,(\nu, t, \beta)$ for some values of t and β.

situation; the value of r probably is close to 1.

In order to find a good value for r, regression (for each combination of t and β) is performed for values of r ranging from -20 to 20 in steps of 0.01 (0 is excluded), resulting in a regression equation $\hat{k}_s(r, \nu, t, \beta) = g_0(r, t, \beta) + g_1(r, t, \beta)\nu^r$. We need to perform these regressions for every value of r because the form in (B.1) does not allow to estimate values for γ_0, γ_1, and r in one regression step. Further, we cannot decide on a good value for r a priori, hence we use a greedy approach to find it. The sum of squared errors (SSE) is determined for each value of r according to

$$SSE(r) = \sum_{\nu \in \mathcal{V}} \sum_{t \in \mathcal{T}} \sum_{\beta \in \mathcal{B}} (k_s(\nu, t, \beta) - \hat{k}_s(r, \nu, t, \beta))^2,$$

where $\mathcal{V}$, $\mathcal{T}$ and $\mathcal{B}$ are the sets of values for ν, t and β (see also (3.12)). The value of r is chosen such that the SSE is minimized. Figure B.2 shows the SSE depending on r. The value of r for which the SSE is lowest, is $\hat{r} = 0.90$. The estimated coefficients

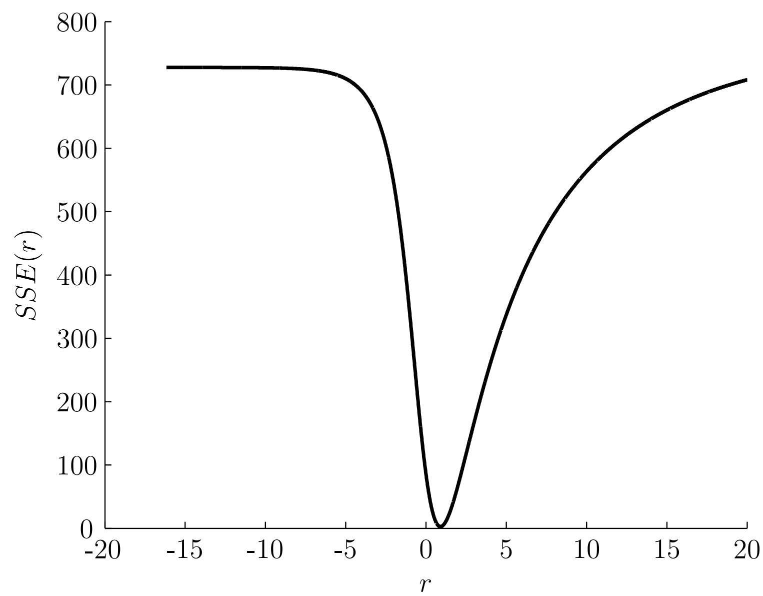


Figure B.2: $SSE(r)$ depending on the exponent r.

g_0 and g_1 for every combination of t and β are listed in Tables B.1 and B.2.

	$\beta = 0.90$	$\beta = 0.91$	$\beta = 0.92$	$\beta = 0.93$	$\beta = 0.94$	$\beta = 0.95$	$\beta = 0.96$	$\beta = 0.97$	$\beta = 0.98$	$\beta = 0.99$
$t = 2$	−1.252	−0.835	−0.687	−0.560	−0.528	−0.520	−0.532	−0.481	−0.480	−0.483
$t = 3$	0.435	0.054	−0.043	−0.100	−0.144	−0.159	−0.154	−0.171	−0.179	−0.177
$t = 4$	0.273	0.077	−0.005	−0.041	−0.057	−0.083	−0.080	−0.077	−0.087	−0.085
$t = 5$	0.199	0.068	0.004	−0.010	−0.021	−0.033	−0.040	−0.042	−0.044	−0.043
$t = 6$	0.157	0.063	0.023	0.010	−0.007	−0.012	−0.017	−0.021	−0.025	−0.023
$t = 7$	0.148	0.059	0.034	0.023	−0.001	−0.002	−0.015	−0.014	−0.014	−0.013
$t = 8$	0.149	0.067	0.035	0.025	0.009	0.000	−0.004	−0.007	−0.011	−0.010
$t = 10$	0.127	0.065	0.040	0.015	0.012	0.001	0.002	−0.008	−0.007	−0.005
$t = 12$	0.093	0.056	0.034	0.022	0.002	−0.002	0.000	−0.007	−0.005	−0.001
$t = 15$	0.077	0.047	0.024	0.012	0.009	0.008	0.004	−0.003	−0.003	−0.006
$t = 20$	0.060	0.034	0.021	0.015	0.004	0.006	−0.001	−0.005	−0.007	−0.001

Table B.1: Estimated values for γ_0.

	$\beta = 0.90$	$\beta = 0.91$	$\beta = 0.92$	$\beta = 0.93$	$\beta = 0.94$	$\beta = 0.95$	$\beta = 0.96$	$\beta = 0.97$	$\beta = 0.98$	$\beta = 0.99$
$t = 2$	28.812	14.135	9.298	6.758	5.299	4.421	3.815	3.209	2.856	2.564
$t = 3$	4.323	2.963	2.238	1.851	1.583	1.356	1.165	1.065	0.965	0.859
$t = 4$	2.082	1.488	1.168	0.975	0.810	0.717	0.619	0.535	0.481	0.422
$t = 5$	1.309	0.922	0.748	0.590	0.492	0.424	0.370	0.324	0.284	0.249
$t = 6$	0.913	0.626	0.500	0.394	0.339	0.284	0.242	0.209	0.185	0.153
$t = 7$	0.659	0.478	0.356	0.281	0.253	0.206	0.179	0.158	0.128	0.105
$t = 8$	0.472	0.352	0.277	0.212	0.187	0.158	0.133	0.115	0.099	0.080
$t = 10$	0.306	0.223	0.178	0.152	0.117	0.114	0.082	0.087	0.072	0.053
$t = 12$	0.252	0.164	0.125	0.103	0.108	0.090	0.070	0.067	0.058	0.032
$t = 15$	0.184	0.115	0.096	0.087	0.065	0.052	0.042	0.043	0.038	0.030
$t = 20$	0.117	0.085	0.063	0.047	0.053	0.034	0.036	0.037	0.031	0.017

Table B.2: Estimated values for γ_1.

So, for example, if we know that $t = 4$ and $\beta = 0.90$, the correction is estimated to be

$$0.273 + 2.082\nu^{0.90}.$$

Note that, although the results in both tables are rounded to three decimals, we have used the unrounded results for the following step.

Step 2 In this step we will regress t on the estimated coefficients of Step 1 for every value of β. We assume the relations

$$g_0(r_0, \beta) = \gamma_{00}(r_0, \beta) + \gamma_{01}(r_0, \beta)t^{r_0} + \varepsilon,$$
$$g_1(r_1, \beta) = \gamma_{10}(r_0, \beta) + \gamma_{11}(r_0, \beta)t^{r_1} + \varepsilon.$$

If we consider $\beta = 0.90$, then we take all g_0 for which $\beta = 0.90$ and regress t^{r_0} on those values. The value of r_0 is chosen in the same way as in Step 1, i.e., we find the value of r for which the SSE is minimized. This value is $\hat{r}_0 = -9.17$. In the same way we estimate γ_{10}, γ_{11}, and the best value for r_1, which is $\hat{r}_1 = -4.19$. The values of g_{00}, g_{01}, g_{10}, and g_{11} are listed in Table B.3. If we know that $\beta = 0.90$, the correction, depending on ν and t, is estimated to be

$$(-0.035 - 259.665t^{-9.17}) + (0.124 + 45.646t^{-4.19})\nu^{0.90}.$$

Again, the results in this table are rounded to three decimals; we have used the unrounded results for the final step.

Step 3 In the last step β is regressed on the four sets of coefficients found in step 2. Actually, we have chosen to use $(1 - \beta)$ as an independent variable, since this is

	g_{00}	g_{01}	g_{10}	g_{11}
$\beta = 0.90$	-0.035	-259.665	0.124	45.646
$\beta = 0.91$	-0.037	-257.118	0.150	50.630
$\beta = 0.92$	-0.034	-259.279	0.170	56.782
$\beta = 0.93$	-0.029	-291.183	0.183	67.574
$\beta = 0.94$	-0.026	-286.663	0.215	78.252
$\beta = 0.95$	-0.018	-295.662	0.248	93.834
$\beta = 0.96$	-0.001	-323.357	0.277	119.946
$\beta = 0.97$	0.019	-406.928	0.314	165.327
$\beta = 0.98$	0.061	-516.219	0.351	252.385
$\beta = 0.99$	0.175	-817.650	0.288	517.293

Table B.3: Estimated values for γ_{00}, γ_{01}, γ_{10}, and γ_{11}.

used in the determination of the safety factor $\hat{c}^{\tau}_{\beta}$; see (3.11). The following relations are assumed:

$$\begin{aligned}
g_{00}(r_{00}, \beta) &= \gamma_{000}(r_{00}) + \gamma_{001}(r_{00})(1-\beta)^{r_{00}} + \varepsilon, \\
g_{01}(r_{01}, \beta) &= \gamma_{010}(r_{01}) + \gamma_{011}(r_{01})(1-\beta)^{r_{01}} + \varepsilon, \\
g_{10}(r_{10}, \beta) &= \gamma_{100}(r_{10}) + \gamma_{101}(r_{10})(1-\beta)^{r_{10}} + \varepsilon, \\
g_{11}(r_{11}, \beta) &= \gamma_{110}(r_{11}) + \gamma_{111}(r_{11})(1-\beta)^{r_{11}} + \varepsilon.
\end{aligned}$$

The coefficients and values of r_{00}, r_{01}, r_{10}, and r_{11} are estimated in the same way as before. The best values for the powers are $\hat{r}_{00} = -0.95$, $\hat{r}_{01} = -1.00$, $\hat{r}_{10} = 1.41$ and $\hat{r}_{11} = -1.03$; the obtained estimates $g_{000}, g_{001}, \ldots, g_{111}$ are in Table B.4. The final

g_{000}	g_{001}	g_{010}	g_{011}	g_{100}	g_{101}	g_{110}	g_{111}
-0.0669	0.00305	-185.124	-6.359	0.335	-5.671	-3.481	4.541

Table B.4: The estimates for the coefficients $\gamma_{000}, \ldots, \gamma_{111}$.

formula for estimating the correction needed is now given by

$$\begin{aligned}
&\hat{k}_s(\nu, t, \beta) = \\
&\Big[\left(-0.0669 + 0.00305(1-\beta)^{-0.95}\right) + \left(-185.124 - 6.359(1-\beta)^{-1.00}\right) t^{-9.17} \Big] \\
&\quad + \Big[\left(0.335 - 5.671(1-\beta)^{1.41}\right) + \left(-3.481 + 4.541(1-\beta)^{-1.03}\right) t^{-4.19} \Big] \nu^{0.90}.
\end{aligned}$$

□

Appendix C

Derivations Chapter 3

C.1 Independence of achieved performance of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$

This appendix shows that the achieved performance depends on the quotient of σ and μ (and thus on the coefficient of variation ν), but that it is independent of μ and σ separately. Three order-up-to levels, $S(\bar{d}_t, s_t\tau, c_\beta^\tau)$, $S(\bar{d}_t, s_t\tau, \hat{c}_\beta^\tau)$ and $S(\bar{d}_t + \hat{\kappa}_i(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}_\beta^\tau)$, are considered. The first is discussed in Section 3.3.2, the second in Section 3.3.3 and the third in Section 3.3.4. Remember that $(x)^+ = \max(0, x)$.

C.1.1 $\boldsymbol{S(\bar{d}_t, s_t\tau, c_\beta^\tau)}$

The service achieved in simulation is denoted by $\hat{\beta}(\boldsymbol{S_n})$, where $\boldsymbol{S_n}$ is the vector of the n order-up-to levels determined for the n samples. It is defined as

$$\hat{\beta}(\boldsymbol{S_n}) = 1 - \frac{\sum_{j=1}^{n}(d_j - S_j)^+}{\sum_{j=1}^{n} d_j}.$$

In this definition d_j denotes the observation that is used to check the order-up-to level obtained in the j-th simulation run (n runs in total). $S_j = S(\bar{d}_{tj}, s_{tj}\tau, c_\beta^\tau)$ is the order-up-to level determined in the j-th simulation, where $\bar{d}_{tj}$ and s_{tj} are defined as

$$\bar{d}_{tj} = \tfrac{1}{t}\sum_{i=1}^{t} d_{ij} \qquad \text{and} \qquad s_{tj} = \sqrt{\tfrac{1}{t-1}\sum_{i=1}^{t}\left(d_{ij} - \bar{d}_{tj}\right)^2}.$$

In the above $d_{1j}, \ldots, d_{tj}$ are the t demand observations that are used in the j-th simulation run to determine the order-up-to level.

Since $d_j \sim N(\mu, \sigma^2)$, $d_j^* = \frac{d_j - \mu}{\sigma} \sim N(0,1)$ and hence d_j^* is indeed independent of μ and σ. Note that the same holds for $d_{ij}^* = (d_{ij} - \mu)/\sigma$. Also $S_j^* = (S_j - \mu)/\sigma$ is independent of μ and σ:

$$\begin{aligned} S_j^* &= \frac{\frac{1}{t}\sum_{i=1}^t d_{ij} + s_{tj} c_\beta^\tau \tau - \mu}{\sigma} = \frac{1}{t}\sum_{i=1}^t \left(\left(\frac{d_{ij} - \mu}{\sigma} \right) + \left(\frac{s_{tj}}{\sigma} \right) c_\beta^\tau \tau \right) \\ &= \frac{1}{t}\sum_{i=1}^t d_{ij}^* + s_{tj}^* c_\beta^\tau \tau. \end{aligned}$$

In the above $s_{tj}^* = s_{tj}/\sigma$. It suffices to show that s_{tj}^* is independent of μ and σ:

$$\begin{aligned} s_{tj}^* &= \frac{s_{tj}}{\sigma} = \frac{\sqrt{\frac{1}{t-1}\sum_{i=1}^t \left(d_{ij} - \frac{1}{t}\sum_{i=1}^t d_{ij} \right)^2}}{\sigma} \\ &= \sqrt{\frac{1}{t-1}\sum_{i=1}^t \left(\frac{d_{ij} - \mu}{\sigma} - \frac{1}{t}\sum_{i=1}^t \frac{d_{ij} - \mu}{\sigma} \right)^2} = \sqrt{\frac{1}{t-1}\sum_{i=1}^t \left(d_{ij}^* - \frac{1}{t}\sum_{i=1}^t d_{ij}^* \right)^2}. \end{aligned}$$

Note that $d_j = d_j^*\sigma + \mu$ and $S_j = S_j^*\sigma + \mu$. If these are substituted in the definition for the simulated performance, we obtain

$$\begin{aligned} \hat{\beta}\left(\boldsymbol{S_n}\right) &= 1 - \frac{\frac{1}{n}\sum_{j=1}^n \left(d_j^*\sigma + \mu - \left(S_j^*\sigma + \mu \right) \right)^+}{\frac{1}{n}\sum_{j=1}^n d_j^*\sigma + \mu} \\ &= 1 - \frac{\frac{1}{n}\sum_{j=1}^n \sigma \left(d_j^* - S_j^* \right)^+}{\frac{1}{n}\sum_{j=1}^n d_j^*\sigma + \mu} = 1 - \frac{\frac{1}{n}\sum_{j=1}^n \left(d_j^* - S_j^* \right)^+}{\frac{1}{n}\sum_{j=1}^n d_j^* + \nu^{-1}}. \end{aligned}$$

Thus the performance only consists of terms independent of μ and σ. It does, however, clearly depend on ν: directly through ν^{-1} in the denominator and indirectly through c_β^τ in S_j^*.

C.1.2 $\boldsymbol{S(\bar{d}_t, s_t\tau, \hat{c}_\beta^\tau)}$

The only thing that changes with respect to the previous section is that now ν has to be estimated (else $\hat{c}_\beta^\tau$ cannot be found). Using the same line of reasoning results in $s_t^* = s_t/\sigma$ and $\bar{d}_t^* = (\bar{d}_t - \mu)/\sigma$ being both independent of μ and σ. So $\hat{\nu}_t = s_t/\bar{d}_t$ is independent of μ and σ:

$$\hat{\nu}_t = \frac{s_t}{\bar{d}_t} = \frac{s_t^*\sigma}{\bar{d}_t^*\sigma + \mu} = \frac{s_t^*}{\bar{d}_t^* + \nu^{-1}}.$$

As stated before $\hat{c}^{\tau}_{\beta}$ is defined as

$$\hat{c}^{\tau}_{\beta} = \begin{cases} G^{-1}\left(\frac{1-\beta}{\hat{\nu}_t \tau}\right) & \text{if } \hat{\nu}_t > 0, \\ -\frac{1}{\hat{\nu}_t \tau} & \text{otherwise.} \end{cases}$$

Thus both for $\hat{\nu}_t > 0$ and $\hat{\nu}_t \leq 0$, $\hat{c}^{\tau}_{\beta}$ purely depends on terms independent of μ and σ. Hence, $\hat{c}^{\tau}_{\beta}$ is independent of μ and σ and it is already shown that the other terms in $(S(\bar{d}_t, s_t\tau, \hat{c}^{\tau}_{\beta}) - \mu)/\sigma$ are. Hence the performance will again be independent of μ and σ.

C.1.3 $S(\bar{d}_t + \hat{\kappa}_i\,(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau, \hat{c}^{\tau}_{\beta})$

So again $(S(\bar{d}_t + \hat{\kappa}_i\,(\hat{\nu}_t, t, \beta)\, s_t, s_t\tau\hat{c}^{\tau}_{\beta}) - \mu)/\sigma$ $(i \in \{\sigma, s\})$ should be independent of μ and σ. This order-up-to level can be rewritten as

$$S(\bar{d}_t + \hat{\kappa}_i\,(\hat{\nu}_t, t, \beta)\, s_t, \hat{c}^{\tau}_{\beta}) = S(\bar{d}_t, s_t\tau, \hat{c}^{\tau}_{\beta}) + \hat{\kappa}_i\,(\hat{\nu}_t, t, \beta)\, s_t.$$

Hence it suffices to show that $\hat{\kappa}_i\,(\hat{\nu}_t, t, \beta)\, s_t/\sigma$ is independent of μ and σ. This is easily seen, if one realizes that $\hat{\nu}_t$ and s_t/σ are independent of μ and σ. Thus the performance of this order-up-to level is independent of μ and σ.

Appendix D

Derivations Chapter 4

D.1 The attained service level in simulation

n	: Number of simulations;
i	: Simulation run ($i = 1, \ldots, n$);
d_{iR}	: Demand during review in ith run;
d_{iL}	: Demand during lead time in ith run;
s_i	: Estimated order-up-to level in ith run;
$\mathrm{I}(c)$	: Indicator function: 1 if c is true and 0 otherwise.

In case of a P_1 criterion the attained service level ($\hat{\alpha}_j$, $j \in \{1, 2, 3, 4\}$) is determined by

$$\hat{\alpha}_j = \frac{\sum_{i=1}^{n} \mathrm{I}(s_i > d_{iR} + d_{iL})}{n}.$$

In case of a P_2 criterion the attained service level ($\hat{\beta}_j$, $j \in \{0, 1, 2, 3, 4\}$) is determined by

$$\hat{\beta}_j = 1 - \frac{\sum_{i=1}^{n} \left[(d_{iR} + d_{iL} - s_i)^+ - (d_{iL} - s_i)^+\right]}{\sum_{i=1}^{n} d_{iR}}.$$

D.2 Equality order-up-to levels under P_1 and P_2 if demand is exponential

Remember that in case of the P_1 criterion the order-up-to level S satisfied the equality $\mathbb{P}\left(D_{1+L} > S\right) = 1-\alpha$ or $\mathbb{P}\left(D^*_{1+L} > S^*\right) = 1-\alpha$, where $S^* = \frac{1}{\theta}S$. Let us now consider the service equality in case of a P_2 criterion. Choose S such that it satisfies

$$\begin{aligned}
&\mathbb{E}\left[(D_{1+L}-S)^+\right]-\mathbb{E}\left[(D_L-S)^+\right]=(1-\beta)\mathbb{E}\left[D\right]\\
\Leftrightarrow\qquad &\theta\left(\mathbb{E}\left[(D^*_{1+L}-S^*)^+\right]-\mathbb{E}\left[(D^*_L-S^*)^+\right]\right)=(1-\beta)\theta\\
\Leftrightarrow\qquad &\mathbb{E}\left[(D^*_{1+L}-S^*)^+\right]-\mathbb{E}\left[(D^*_L-S^*)^+\right]=1-\beta. \qquad (D.1)
\end{aligned}$$

Note that, analogously to the P_1 case, $D^*_\ell = \frac{1}{\theta}D_\ell \sim \Gamma(\ell, 1)$ and $S^* = \frac{1}{\theta}S$. Since it is assumed that the desired service levels under the P_1 and P_2 policy are chosen to be equal, we have $\alpha = \beta$. Hence, the right hand sides of both service equations are equal, so if the left hand sides of the service equations are equal as well, then $S^* = S^*$ or $S = S$. So consider the left hand side of (D.1). Note that $rf_{1+r,1}(x) = xf_{r,1}(x)$ and $F_{1+r,1}(x) = F_{r,1}(x) - f_{1+r,1}(x)$. We obtain that

$$\begin{aligned}
&\mathbb{E}\left[(D^*_{1+L}-S^*)^+\right]-\mathbb{E}\left[(D^*_L-S^*)^+\right]\\
&\quad=\mathcal{L}_{1+L,1}(S^*)-\mathcal{L}_{L,1}(S^*)\\
&\quad=(1+L)[1-F_{2+L,1}(S^*)]-S^*[1-F_{1+L}(S^*)]-\mathcal{L}_{L,1}(S^*)\\
&\quad=(1+L)[1-F_{1+L,1}(S^*)+f_{2+L}(S^*)]\\
&\qquad-S^*[1-F_L(S^*)+f_{1+L}(S^*)]-\mathcal{L}_{L,1}(S^*)\\
&\quad=(1+L)[1-F_{1+L,1}(S^*)]-S^*[1-F_L(S^*)] \qquad (D.2)\\
&\qquad+(1+L)f_{2+L}(S^*)-S^*f_{1+L}(S^*)-\mathcal{L}_{L,1}(S^*)\\
&\quad=1-F_{1+L,1}(S^*)+\mathcal{L}_{L,1}(S^*)+S^*f_{1+L}(S^*)-S^*f_{1+L}(S^*)-\mathcal{L}_{L,1}(S^*)\\
&\quad=1-F_{1+L,1}(S^*)=\mathbb{P}\left(D^*_{1+L}>S^*\right).
\end{aligned}$$

Thus, the left hand sides of the service equations are equal as well and therefore the order-up-to levels under the P_1 and P_2 criterion will be equal if demand is exponentially distributed.

D.3 Simulation results

The tables below display the results of the calculations and simulations for figures and tables with extreme values in Chapter 4.

		$\alpha = 0.90$			$\alpha = 0.95$			$\alpha = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{3}$	0.8158	0.8572	0.8714	0.8681	0.9094	0.9232	0.9269	0.9621	0.9726
	$\frac{44}{13}$	0.8541	0.8761	0.8839	0.9102	0.9299	0.9366	0.9682	0.9801	0.9837
	$\frac{165}{17}$	0.8624	0.8804	0.8868	0.9184	0.9340	0.9393	0.9743	0.9828	0.9853
1	$\frac{1}{3}$	0.7696	0.8304	0.8527	0.8238	0.8850	0.9066	0.8910	0.9453	0.9623
	$\frac{44}{13}$	0.8261	0.8600	0.8726	0.8846	0.9161	0.9272	0.9516	0.9727	0.9790
	$\frac{165}{17}$	0.8370	0.8659	0.8766	0.8956	0.9218	0.9310	0.9605	0.9766	0.9815
4	$\frac{1}{3}$	0.6935	0.7783	0.8138	0.7470	0.8353	0.8708	0.8220	0.9070	0.9374
	$\frac{44}{13}$	0.7702	0.8229	0.8452	0.8293	0.8821	0.9031	0.9083	0.9510	0.9653
	$\frac{165}{17}$	0.7841	0.8311	0.8510	0.8437	0.8903	0.9088	0.9219	0.9576	0.9695
6	$\frac{1}{3}$	0.6635	0.7542	0.7944	0.7153	0.8113	0.8523	0.7910	0.8866	0.9232
	$\frac{44}{13}$	0.7450	0.8037	0.8300	0.8028	0.8634	0.8891	0.8844	0.9373	0.9563
	$\frac{165}{17}$	0.7596	0.8127	0.8366	0.8181	0.8726	0.8955	0.8996	0.9452	0.9613

Table D.1: Attained service level for Figure 4.2.

L	ρ	$\alpha = 0.90$ $t=4$	$t=8$	$t=12$	$\alpha = 0.95$ $t=4$	$t=8$	$t=12$	$\alpha = 0.99$ $t=4$	$t=8$	$t=12$
0	$\frac{1}{2}$	0.7991	0.8447	0.8606	0.8445	0.8933	0.9112	0.9000	0.9485	0.9638
	1	0.8054	0.8494	0.8662	0.8567	0.9023	0.9171	0.9169	0.9564	0.9681
	$\frac{44}{13}$	0.8196	0.8594	0.8730	0.8718	0.9103	0.9236	0.9299	0.9649	0.9734
	6	0.8238	0.8615	0.8731	0.8746	0.9114	0.9262	0.9312	0.9657	0.9749
	9	0.8264	0.8620	0.8756	0.8781	0.9142	0.9264	0.9333	0.9659	0.9761
1	$\frac{1}{2}$	0.7483	0.8146	0.8400	0.7959	0.8676	0.8924	0.8592	0.9270	0.9499
	1	0.7670	0.8256	0.8507	0.8169	0.8826	0.9010	0.8862	0.9413	0.9596
	$\frac{44}{13}$	0.7917	0.8410	0.8604	0.8437	0.8959	0.9138	0.9082	0.9533	0.9675
	6	0.7967	0.8468	0.8637	0.8508	0.8992	0.9173	0.9120	0.9568	0.9692
	9	0.8015	0.8480	0.8650	0.8533	0.9030	0.9178	0.9156	0.9581	0.9705
$4\frac{1}{3}$	$\frac{1}{2}$	0.6685	0.7554	0.7975	0.7142	0.8118	0.8525	0.7855	0.8822	0.9200
	1	0.6993	0.7776	0.8097	0.7477	0.8327	0.8688	0.8216	0.9050	0.9353
	$\frac{44}{13}$	0.7357	0.8016	0.8295	0.7869	0.8586	0.8850	0.8586	0.9255	0.9487
	6	0.7419	0.8088	0.8346	0.7946	0.8645	0.8878	0.8685	0.9307	0.9526
	9	0.7484	0.8101	0.8365	0.7993	0.8671	0.8912	0.8714	0.9339	0.9551
6	$\frac{1}{2}$	0.6449	0.7405	0.7818	0.6911	0.7894	0.8370	0.7579	0.8653	0.9072
	1	0.6785	0.7611	0.7982	0.7291	0.8159	0.8570	0.8014	0.8912	0.9255
	$\frac{44}{13}$	0.7165	0.7872	0.8153	0.7656	0.8411	0.8727	0.8378	0.9152	0.9411
	6	0.7255	0.7942	0.8203	0.7753	0.8490	0.8794	0.8484	0.9192	0.9448
	9	0.7332	0.7970	0.8243	0.7834	0.8527	0.8819	0.8518	0.9246	0.9470
9	$\frac{1}{2}$	0.6144	0.7106	0.7599	0.6575	0.7658	0.8132	0.7244	0.8375	0.8877
	1	0.6497	0.7353	0.7761	0.6970	0.7880	0.8332	0.7675	0.8677	0.9091
	$\frac{44}{13}$	0.6924	0.7634	0.7992	0.7367	0.8184	0.8552	0.8110	0.8952	0.9261
	6	0.7030	0.7698	0.8062	0.7483	0.8272	0.8609	0.8210	0.9022	0.9321
	9	0.7063	0.7745	0.8073	0.7563	0.8302	0.8633	0.8276	0.9050	0.9339

Table D.2: Attained service level for Figure 4.3.

		$\alpha = 0.90$			$\alpha = 0.95$			$\alpha = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.8515	0.8740	0.8811	0.8989	0.9211	0.9302	0.9534	0.9711	0.9769
	1	0.8632	0.8802	0.8870	0.9141	0.9305	0.9360	0.9651	0.9774	0.9806
	$\frac{44}{13}$	0.8768	0.8905	0.8942	0.9260	0.9383	0.9425	0.9731	0.9833	0.9855
	6	0.8802	0.8919	0.8941	0.9280	0.9397	0.9449	0.9742	0.9837	0.9865
	9	0.8824	0.8929	0.8968	0.9312	0.9419	0.9453	0.9746	0.9843	0.9865
1	$\frac{1}{2}$	0.8032	0.8454	0.8608	0.8553	0.8988	0.9127	0.9249	0.9549	0.9661
	1	0.8266	0.8570	0.8721	0.8801	0.9134	0.9218	0.9471	0.9677	0.9747
	$\frac{44}{13}$	0.8517	0.8730	0.8819	0.9037	0.9265	0.9338	0.9624	0.9761	0.9812
	6	0.8571	0.8786	0.8857	0.9099	0.9288	0.9370	0.9641	0.9779	0.9823
	9	0.8593	0.8799	0.8864	0.9114	0.9329	0.9375	0.9655	0.9792	0.9834
$4\frac{1}{3}$	$\frac{1}{2}$	0.7220	0.7860	0.8190	0.7787	0.8467	0.8759	0.8678	0.9197	0.9421
	1	0.7568	0.8095	0.8323	0.8162	0.8679	0.8919	0.9021	0.9403	0.9561
	$\frac{44}{13}$	0.7940	0.8350	0.8524	0.8532	0.8929	0.9083	0.9305	0.9570	0.9679
	6	0.8005	0.8420	0.8567	0.8613	0.8990	0.9108	0.9374	0.9616	0.9699
	9	0.8070	0.8429	0.8590	0.8641	0.9010	0.9129	0.9399	0.9634	0.9729
6	$\frac{1}{2}$	0.6964	0.7705	0.8030	0.7554	0.8253	0.8605	0.8445	0.9053	0.9319
	1	0.7328	0.7933	0.8201	0.7960	0.8523	0.8800	0.8861	0.9302	0.9476
	$\frac{44}{13}$	0.7736	0.8196	0.8377	0.8344	0.8772	0.8975	0.9160	0.9504	0.9617
	6	0.7834	0.8268	0.8430	0.8426	0.8842	0.9026	0.9236	0.9532	0.9652
	9	0.7913	0.8295	0.8466	0.8491	0.8880	0.9053	0.9253	0.9571	0.9658
9	$\frac{1}{2}$	0.6624	0.7400	0.7807	0.7197	0.8005	0.8374	0.8134	0.8811	0.9145
	1	0.7032	0.7660	0.7986	0.7623	0.8251	0.8581	0.8571	0.9087	0.9345
	$\frac{44}{13}$	0.7492	0.7951	0.8208	0.8032	0.8541	0.8795	0.8949	0.9335	0.9494
	6	0.7583	0.8027	0.8286	0.8172	0.8628	0.8847	0.9025	0.9403	0.9549
	9	0.7639	0.8066	0.8299	0.8240	0.8656	0.8880	0.9086	0.9422	0.9561

Table D.3: Attained service level for Figure 4.4.

		$\alpha = 0.90$			$\alpha = 0.95$			$\alpha = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.8826	0.8935	0.8961	0.9290	0.9414	0.9446	0.9756	0.9836	0.9861
	1	0.8883	0.8949	0.8976	0.9373	0.9449	0.9467	0.9804	0.9857	0.9870
	$\frac{44}{13}$	0.8924	0.8982	0.8995	0.9417	0.9464	0.9486	0.9841	0.9881	0.9886
	6	0.8938	0.8975	0.8977	0.9419	0.9461	0.9493	0.9858	0.9881	0.9893
	9	0.8956	0.8977	0.8998	0.9464	0.9485	0.9493	0.9877	0.9887	0.9893
1	$\frac{1}{2}$	0.8654	0.8876	0.8924	0.9149	0.9365	0.9414	0.9676	0.9789	0.9840
	1	0.8767	0.8913	0.8974	0.9247	0.9416	0.9436	0.9744	0.9833	0.9855
	$\frac{44}{13}$	0.8868	0.8952	0.8991	0.9357	0.9467	0.9475	0.9806	0.9861	0.9881
	6	0.8898	0.8984	0.9008	0.9393	0.9457	0.9498	0.9836	0.9874	0.9885
	9	0.8915	0.8995	0.9002	0.9414	0.9487	0.9498	0.9856	0.9884	0.9896
$4\frac{1}{3}$	$\frac{1}{2}$	0.8434	0.8730	0.8878	0.8953	0.9254	0.9347	0.9543	0.9743	0.9796
	1	0.8616	0.8834	0.8898	0.9103	0.9301	0.9399	0.9634	0.9789	0.9827
	$\frac{44}{13}$	0.8746	0.8911	0.8953	0.9226	0.9406	0.9428	0.9721	0.9830	0.9866
	6	0.8785	0.8957	0.8979	0.9274	0.9418	0.9443	0.9767	0.9848	0.9861
	9	0.8840	0.8967	0.9001	0.9303	0.9442	0.9458	0.9791	0.9860	0.9882
6	$\frac{1}{2}$	0.8399	0.8757	0.8849	0.8916	0.9222	0.9343	0.9508	0.9722	0.9789
	1	0.8575	0.8831	0.8894	0.9081	0.9305	0.9393	0.9615	0.9776	0.9822
	$\frac{44}{13}$	0.8733	0.8910	0.8941	0.9203	0.9378	0.9434	0.9696	0.9829	0.9850
	6	0.8782	0.8943	0.8975	0.9245	0.9399	0.9447	0.9737	0.9840	0.9863
	9	0.8865	0.8970	0.8992	0.9309	0.9434	0.9462	0.9760	0.9857	0.9870
9	$\frac{1}{2}$	0.8387	0.8734	0.8854	0.8899	0.9221	0.9335	0.9479	0.9709	0.9778
	1	0.8587	0.8818	0.8895	0.9055	0.9295	0.9368	0.9587	0.9763	0.9819
	$\frac{44}{13}$	0.8747	0.8905	0.8955	0.9161	0.9368	0.9421	0.9658	0.9805	0.9844
	6	0.8809	0.8957	0.9008	0.9228	0.9414	0.9436	0.9683	0.9842	0.9863
	9	0.8872	0.8990	0.9034	0.9292	0.9436	0.9467	0.9730	0.9850	0.9877

Table D.4: Attained service level for Figure 4.6.

		$\alpha = 0.90$			$\alpha = 0.95$			$\alpha = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.9014	0.9006	0.8992	0.9490	0.9502	0.9493	0.9895	0.9900	0.9897
	1	0.9002	0.8988	0.8994	0.9504	0.9499	0.9492	0.9901	0.9897	0.9893
	$\frac{44}{13}$	0.8988	0.8999	0.9005	0.9486	0.9487	0.9497	0.9891	0.9899	0.9897
	6	0.8979	0.8991	0.8985	0.9478	0.9479	0.9502	0.9895	0.9893	0.9900
	9	0.8991	0.8985	0.9000	0.9504	0.9498	0.9501	0.9902	0.9898	0.9898
1	$\frac{1}{2}$	0.8998	0.9009	0.8996	0.9482	0.9507	0.9497	0.9895	0.9897	0.9897
	1	0.8991	0.8993	0.9016	0.9492	0.9501	0.9484	0.9901	0.9894	0.9893
	$\frac{44}{13}$	0.8993	0.8994	0.9009	0.9496	0.9511	0.9499	0.9894	0.9894	0.9896
	6	0.9001	0.9011	0.9022	0.9502	0.9496	0.9515	0.9900	0.9897	0.9898
	9	0.9000	0.9017	0.9016	0.9506	0.9515	0.9516	0.9903	0.9903	0.9908
$4\frac{1}{3}$	$\frac{1}{2}$	0.8992	0.8984	0.9023	0.9494	0.9500	0.9501	0.9895	0.9898	0.9900
	1	0.8998	0.8988	0.8982	0.9488	0.9473	0.9494	0.9904	0.9895	0.9894
	$\frac{44}{13}$	0.8972	0.9000	0.9003	0.9473	0.9490	0.9476	0.9889	0.9889	0.9897
	6	0.8972	0.9013	0.9016	0.9475	0.9486	0.9479	0.9897	0.9893	0.9890
	9	0.9005	0.9022	0.9024	0.9473	0.9503	0.9492	0.9901	0.9898	0.9903
6	$\frac{1}{2}$	0.9044	0.9031	0.9029	0.9509	0.9508	0.9522	0.9895	0.9903	0.9903
	1	0.9023	0.9008	0.8997	0.9516	0.9487	0.9502	0.9909	0.9900	0.9896
	$\frac{44}{13}$	0.9007	0.9008	0.8996	0.9484	0.9484	0.9488	0.9891	0.9897	0.9893
	6	0.8992	0.9022	0.9012	0.9483	0.9485	0.9489	0.9893	0.9895	0.9894
	9	0.9047	0.9034	0.9025	0.9504	0.9505	0.9499	0.9892	0.9903	0.9895
9	$\frac{1}{2}$	0.9109	0.9082	0.9077	0.9556	0.9544	0.9545	0.9909	0.9910	0.9906
	1	0.9109	0.9051	0.9033	0.9550	0.9523	0.9500	0.9919	0.9906	0.9909
	$\frac{44}{13}$	0.9079	0.9025	0.9022	0.9498	0.9486	0.9493	0.9899	0.9891	0.9890
	6	0.9063	0.9060	0.9058	0.9505	0.9510	0.9493	0.9886	0.9906	0.9897
	9	0.9086	0.9071	0.9077	0.9526	0.9515	0.9516	0.9893	0.9904	0.9906

Table D.5: All attained service levels for Table 4.10.

		$\beta = 0.90$			$\beta = 0.95$			$\beta = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.7950	0.8463	0.8605	0.8515	0.9009	0.9180	0.9188	0.9589	0.9714
	1	0.8393	0.8663	0.8798	0.8919	0.9215	0.9315	0.9536	0.9731	0.9786
	$\frac{44}{13}$	0.8715	0.8852	0.8910	0.9237	0.9364	0.9420	0.9748	0.9837	0.9857
	6	0.8792	0.8895	0.8926	0.9305	0.9396	0.9443	0.9791	0.9849	0.9870
	9	0.8835	0.8911	0.8942	0.9346	0.9421	0.9451	0.9811	0.9857	0.9873
1	$\frac{1}{2}$	0.7662	0.8264	0.8456	0.8209	0.8855	0.9026	0.8934	0.9448	0.9625
	1	0.8024	0.8485	0.8659	0.8596	0.9048	0.9185	0.9306	0.9622	0.9738
	$\frac{44}{13}$	0.8408	0.8675	0.8789	0.8973	0.9231	0.9319	0.9595	0.9750	0.9808
	6	0.8522	0.8742	0.8835	0.9074	0.9279	0.9359	0.9658	0.9793	0.9829
	9	0.8583	0.8788	0.8855	0.9130	0.9332	0.9379	0.9702	0.9810	0.9841
$4\frac{1}{3}$	$\frac{1}{2}$	0.7029	0.7776	0.8168	0.7558	0.8383	0.8718	0.8313	0.9108	0.9398
	1	0.7339	0.8011	0.8286	0.7938	0.8591	0.8868	0.8729	0.9319	0.9535
	$\frac{44}{13}$	0.7741	0.8256	0.8476	0.8329	0.8858	0.9050	0.9110	0.9515	0.9663
	6	0.7836	0.8341	0.8532	0.8447	0.8932	0.9089	0.9223	0.9578	0.9682
	9	0.7937	0.8383	0.8566	0.8510	0.8972	0.9124	0.9299	0.9616	0.9719
6	$\frac{1}{2}$	0.6778	0.7646	0.8000	0.7331	0.8161	0.8587	0.8098	0.8977	0.9322
	1	0.7121	0.7842	0.8118	0.7699	0.8435	0.8770	0.8536	0.9196	0.9453
	$\frac{44}{13}$	0.7518	0.8104	0.8325	0.8082	0.8690	0.8927	0.8896	0.9422	0.9588
	6	0.7632	0.8179	0.8386	0.8216	0.8770	0.8987	0.9035	0.9478	0.9627
	9	0.7743	0.8225	0.8438	0.8316	0.8818	0.9025	0.9078	0.9526	0.9649
9	$\frac{1}{2}$	0.6491	0.7376	0.7789	0.6987	0.7957	0.8362	0.7794	0.8698	0.9121
	1	0.6823	0.7596	0.7934	0.7381	0.8151	0.8538	0.8190	0.8973	0.9306
	$\frac{44}{13}$	0.7250	0.7828	0.8143	0.7760	0.8426	0.8736	0.8595	0.9225	0.9449
	6	0.7368	0.7926	0.8226	0.7893	0.8533	0.8797	0.8717	0.9309	0.9508
	9	0.7414	0.7967	0.8253	0.7993	0.8566	0.8835	0.8787	0.9335	0.9531

Table D.6: Attained service level for Figure 4.7.

		$\beta = 0.90$			$\beta = 0.95$			$\beta = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.7078	0.7853	0.8154	0.7648	0.8455	0.8759	0.8448	0.9192	0.9441
	1	0.7803	0.8291	0.8526	0.8340	0.8885	0.9076	0.9024	0.9479	0.9614
	$\frac{44}{13}$	0.8457	0.8714	0.8810	0.8941	0.9200	0.9306	0.9475	0.9710	0.9771
	6	0.8629	0.8807	0.8866	0.9094	0.9284	0.9366	0.9573	0.9750	0.9806
	9	0.8723	0.8850	0.8903	0.9187	0.9337	0.9394	0.9636	0.9776	0.9820
1	$\frac{1}{2}$	0.6905	0.7760	0.8047	0.7411	0.8342	0.8642	0.8187	0.9033	0.9334
	1	0.7468	0.8141	0.8404	0.8006	0.8707	0.8928	0.8751	0.9335	0.9539
	$\frac{44}{13}$	0.8080	0.8487	0.8657	0.8596	0.9033	0.9179	0.9219	0.9579	0.9703
	6	0.8272	0.8608	0.8741	0.8789	0.9123	0.9254	0.9351	0.9654	0.9741
	9	0.8395	0.8687	0.8785	0.8892	0.9209	0.9294	0.9445	0.9698	0.9773
$4\frac{1}{3}$	$\frac{1}{2}$	0.6408	0.7334	0.7830	0.6868	0.7924	0.8354	0.7587	0.8662	0.9063
	1	0.6878	0.7693	0.8037	0.7390	0.8246	0.8622	0.8150	0.9001	0.9307
	$\frac{44}{13}$	0.7425	0.8065	0.8334	0.7939	0.8636	0.8891	0.8655	0.9293	0.9514
	6	0.7564	0.8181	0.8415	0.8101	0.8735	0.8950	0.8828	0.9381	0.9564
	9	0.7692	0.8238	0.8466	0.8201	0.8801	0.9002	0.8921	0.9442	0.9615
6	$\frac{1}{2}$	0.6200	0.7217	0.7668	0.6684	0.7698	0.8235	0.7390	0.8511	0.9003
	1	0.6691	0.7545	0.7893	0.7196	0.8115	0.8524	0.7974	0.8869	0.9215
	$\frac{44}{13}$	0.7219	0.7921	0.8182	0.7712	0.8460	0.8766	0.8441	0.9193	0.9435
	6	0.7366	0.8014	0.8271	0.7877	0.8572	0.8851	0.8607	0.9262	0.9494
	9	0.7499	0.8080	0.8331	0.8006	0.8641	0.8900	0.8689	0.9342	0.9526
9	$\frac{1}{2}$	0.5970	0.6987	0.7468	0.6381	0.7515	0.8018	0.7115	0.8237	0.8801
	1	0.6433	0.7305	0.7709	0.6915	0.7834	0.8300	0.7642	0.8632	0.9062
	$\frac{44}{13}$	0.6963	0.7661	0.8012	0.7415	0.8212	0.8582	0.8150	0.8980	0.9283
	6	0.7118	0.7770	0.8123	0.7577	0.8340	0.8656	0.8304	0.9084	0.9362
	9	0.7185	0.7827	0.8142	0.7695	0.8391	0.8708	0.8412	0.9134	0.9398

Table D.7: Attained service level for Figure 4.8.

L	ρ	$\beta = 0.90$ $t=4$	$t=8$	$t=12$	$\beta = 0.95$ $t=4$	$t=8$	$t=12$	$\beta = 0.99$ $t=4$	$t=8$	$t=12$
0	$\frac{1}{2}$	0.7747	0.8206	0.8388	0.8392	0.8822	0.8995	0.9234	0.9522	0.9631
	1	0.8430	0.8616	0.8747	0.8983	0.9203	0.9286	0.9600	0.9724	0.9764
	$\frac{44}{13}$	0.9005	0.9014	0.9015	0.9437	0.9466	0.9486	0.9826	0.9871	0.9876
	6	0.9152	0.9100	0.9070	0.9547	0.9537	0.9538	0.9869	0.9894	0.9899
	9	0.9234	0.9142	0.9106	0.9616	0.9581	0.9563	0.9895	0.9908	0.9908
1	$\frac{1}{2}$	0.7527	0.8091	0.8273	0.8135	0.8698	0.8875	0.9019	0.9388	0.9546
	1	0.8085	0.8472	0.8623	0.8683	0.9034	0.9148	0.9423	0.9626	0.9706
	$\frac{44}{13}$	0.8665	0.8803	0.8870	0.9181	0.9333	0.9373	0.9703	0.9791	0.9829
	6	0.8850	0.8919	0.8952	0.9333	0.9408	0.9444	0.9774	0.9839	0.9857
	9	0.8953	0.8991	0.8994	0.9409	0.9477	0.9477	0.9815	0.9863	0.9879
$4\frac{1}{3}$	$\frac{1}{2}$	0.6963	0.7653	0.8049	0.7551	0.8292	0.8596	0.8475	0.9080	0.9314
	1	0.7457	0.8015	0.8262	0.8090	0.8609	0.8862	0.8971	0.9372	0.9524
	$\frac{44}{13}$	0.8013	0.8391	0.8558	0.8613	0.8980	0.9114	0.9359	0.9601	0.9698
	6	0.8164	0.8508	0.8637	0.8763	0.9067	0.9171	0.9482	0.9668	0.9732
	9	0.8292	0.8564	0.8688	0.8847	0.9130	0.9219	0.9541	0.9707	0.9771
6	$\frac{1}{2}$	0.6732	0.7528	0.7886	0.7355	0.8065	0.8479	0.8295	0.8942	0.9264
	1	0.7252	0.7869	0.8116	0.7875	0.8478	0.8756	0.8833	0.9262	0.9446
	$\frac{44}{13}$	0.7795	0.8246	0.8405	0.8400	0.8817	0.9003	0.9216	0.9533	0.9633
	6	0.7957	0.8340	0.8496	0.8555	0.8919	0.9076	0.9342	0.9585	0.9681
	9	0.8092	0.8409	0.8552	0.8668	0.8986	0.9124	0.9391	0.9639	0.9704
9	$\frac{1}{2}$	0.6472	0.7290	0.7681	0.7019	0.7876	0.8271	0.8026	0.8695	0.9079
	1	0.6963	0.7610	0.7929	0.7581	0.8205	0.8545	0.8540	0.9051	0.9321
	$\frac{44}{13}$	0.7524	0.7978	0.8230	0.8086	0.8578	0.8823	0.8992	0.9356	0.9512
	6	0.7682	0.8093	0.8343	0.8262	0.8702	0.8894	0.9107	0.9449	0.9579
	9	0.7761	0.8152	0.8365	0.8376	0.8750	0.8944	0.9200	0.9490	0.9609

Table D.8: Attained service level for Figure 4.9.

		$\beta = 0.90$			$\beta = 0.95$			$\beta = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.8344	0.8681	0.8789	0.8912	0.9219	0.9316	0.9564	0.9747	0.9802
	1	0.8651	0.8820	0.8931	0.9175	0.9374	0.9441	0.9704	0.9811	0.9841
	$\frac{44}{13}$	0.8862	0.8959	0.8988	0.9363	0.9442	0.9481	0.9806	0.9869	0.9881
	6	0.8887	0.8955	0.8974	0.9398	0.9455	0.9486	0.9827	0.9873	0.9888
	9	0.8885	0.8936	0.8963	0.9420	0.9461	0.9481	0.9841	0.9875	0.9886
1	$\frac{1}{2}$	0.8387	0.8712	0.8789	0.8929	0.9237	0.9307	0.9569	0.9734	0.9802
	1	0.8608	0.8869	0.8937	0.9145	0.9355	0.9413	0.9677	0.9802	0.9840
	$\frac{44}{13}$	0.8805	0.8925	0.8972	0.9311	0.9443	0.9466	0.9771	0.9847	0.9875
	6	0.8855	0.8946	0.8981	0.9374	0.9449	0.9484	0.9807	0.9864	0.9882
	9	0.8873	0.8956	0.8975	0.9394	0.9478	0.9484	0.9828	0.9874	0.9891
$4\frac{1}{3}$	$\frac{1}{2}$	0.8283	0.8638	0.8826	0.8847	0.9191	0.9300	0.9471	0.9710	0.9766
	1	0.8529	0.8780	0.8867	0.9045	0.9268	0.9377	0.9598	0.9774	0.9816
	$\frac{44}{13}$	0.8686	0.8881	0.8939	0.9182	0.9388	0.9429	0.9676	0.9819	0.9861
	6	0.8716	0.8911	0.8947	0.9229	0.9396	0.9434	0.9732	0.9837	0.9860
	9	0.8763	0.8916	0.8962	0.9249	0.9419	0.9447	0.9753	0.9849	0.9878
6	$\frac{1}{2}$	0.8272	0.8645	0.8775	0.8809	0.9140	0.9282	0.9459	0.9692	0.9779
	1	0.8499	0.8770	0.8829	0.9034	0.9268	0.9363	0.9593	0.9759	0.9815
	$\frac{44}{13}$	0.8685	0.8891	0.8914	0.9169	0.9372	0.9428	0.9670	0.9824	0.9852
	6	0.8737	0.8922	0.8953	0.9223	0.9399	0.9444	0.9713	0.9836	0.9863
	9	0.8821	0.8949	0.8972	0.9277	0.9432	0.9464	0.9736	0.9855	0.9871
9	$\frac{1}{2}$	0.8231	0.8620	0.8747	0.8781	0.9118	0.9263	0.9447	0.9669	0.9747
	1	0.8512	0.8741	0.8833	0.9022	0.9261	0.9335	0.9562	0.9751	0.9810
	$\frac{44}{13}$	0.8740	0.8890	0.8934	0.9160	0.9367	0.9431	0.9661	0.9809	0.9849
	6	0.8828	0.8970	0.9014	0.9243	0.9438	0.9453	0.9693	0.9848	0.9872
	9	0.8893	0.9016	0.9039	0.9323	0.9465	0.9495	0.9743	0.9861	0.9885

Table D.9: Attained service level for Figure 4.11.

		$\beta = 0.90$			$\beta = 0.95$			$\beta = 0.99$		
L	ρ	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$	$t = 4$	$t = 8$	$t = 12$
0	$\frac{1}{2}$	0.8988	0.9000	0.8979	0.9488	0.9498	0.9495	0.9892	0.9895	0.9896
	1	0.9004	0.8966	0.9014	0.9494	0.9507	0.9517	0.9897	0.9891	0.9889
	$\frac{44}{13}$	0.8992	0.8999	0.9009	0.9495	0.9489	0.9506	0.9891	0.9899	0.9898
	6	0.8969	0.8979	0.8985	0.9490	0.9485	0.9501	0.9888	0.9894	0.9899
	9	0.8944	0.8952	0.8970	0.9490	0.9482	0.9491	0.9889	0.9892	0.9895
1	$\frac{1}{2}$	0.9012	0.9015	0.8982	0.9494	0.9515	0.9482	0.9899	0.9899	0.9901
	1	0.9003	0.9031	0.9030	0.9518	0.9507	0.9506	0.9905	0.9899	0.9900
	$\frac{44}{13}$	0.9001	0.8990	0.9006	0.9512	0.9510	0.9502	0.9896	0.9893	0.9898
	6	0.9000	0.8989	0.9003	0.9518	0.9498	0.9508	0.9900	0.9897	0.9899
	9	0.8986	0.8989	0.8991	0.9511	0.9515	0.9503	0.9902	0.9899	0.9904
$4\frac{1}{3}$	$\frac{1}{2}$	0.8987	0.9004	0.9033	0.9498	0.9509	0.9496	0.9895	0.9903	0.9894
	1	0.9004	0.8990	0.8989	0.9501	0.9482	0.9499	0.9911	0.9898	0.9897
	$\frac{44}{13}$	0.8970	0.8983	0.8993	0.9486	0.9494	0.9488	0.9891	0.9891	0.9900
	6	0.8951	0.8986	0.8986	0.9475	0.9480	0.9479	0.9896	0.9892	0.9891
	9	0.8961	0.8979	0.8994	0.9460	0.9488	0.9484	0.9894	0.9895	0.9902
6	$\frac{1}{2}$	0.8999	0.8986	0.8993	0.9483	0.9488	0.9489	0.9887	0.9898	0.9902
	1	0.9003	0.8985	0.8960	0.9519	0.9476	0.9488	0.9909	0.9900	0.9897
	$\frac{44}{13}$	0.8998	0.9003	0.8975	0.9492	0.9492	0.9494	0.9900	0.9900	0.9897
	6	0.8986	0.9006	0.8997	0.9490	0.9490	0.9492	0.9895	0.9898	0.9895
	9	0.9030	0.9020	0.9009	0.9503	0.9509	0.9505	0.9895	0.9903	0.9899
9	$\frac{1}{2}$	0.8964	0.8991	0.8996	0.9486	0.9467	0.9494	0.9899	0.9892	0.9887
	1	0.9048	0.8986	0.8975	0.9542	0.9499	0.9475	0.9914	0.9903	0.9904
	$\frac{44}{13}$	0.9082	0.9013	0.9000	0.9517	0.9495	0.9500	0.9912	0.9898	0.9897
	6	0.9092	0.9067	0.9063	0.9529	0.9537	0.9508	0.9905	0.9913	0.9907
	9	0.9118	0.9091	0.9081	0.9564	0.9547	0.9539	0.9910	0.9914	0.9914

Table D.10: All attained service levels for Table 4.17.

OUL	L	$\alpha = 0.90$ $t=4$	$t=8$	$t=12$	$\alpha = 0.95$ $t=4$	$t=8$	$t=12$	$\alpha = 0.99$ $t=4$	$t=8$	$t=12$
Using α										
	0	0.8172	0.8659	0.8709	0.8670	0.9146	0.9232	0.9232	0.9623	0.9704
	1	0.7923	0.8410	0.8381	0.8377	0.8923	0.8973	0.8980	0.9456	0.9529
	4	0.7297	0.7719	0.8106	0.7733	0.8292	0.8642	0.8357	0.8994	0.9280
Using α'										
	0	0.8737	0.8950	0.8917	0.9198	0.9385	0.9417	0.9629	0.9799	0.9818
	1	0.8435	0.8723	0.8607	0.8935	0.9205	0.9188	0.9492	0.9666	0.9659
	4	0.7796	0.8059	0.8309	0.8315	0.8637	0.8831	0.9077	0.9341	0.9499
Using $\hat{k}_\alpha$										
	0	0.8906	0.9044	0.8993	0.9351	0.9493	0.9488	0.9764	0.9838	0.9870
	1	0.8817	0.8972	0.8849	0.9268	0.9409	0.9367	0.9709	0.9794	0.9764
	4	0.8569	0.8649	0.8753	0.9033	0.9156	0.9241	0.9539	0.9691	0.9725

Table D.11: Attained service level for Figure 4.13.

OUL	L	$\beta = 0.90$ $t=4$	$t=8$	$t=12$	$\beta = 0.95$ $t=4$	$t=8$	$t=12$	$\beta = 0.99$ $t=4$	$t=8$	$t=12$
Using β										
	0	0.8163	0.8689	0.8739	0.8640	0.9192	0.9290	0.9150	0.9634	0.9762
	1	0.7966	0.8699	0.8545	0.8392	0.9194	0.9048	0.8874	0.9654	0.9565
	4	0.7705	0.7950	0.8131	0.8244	0.8592	0.8684	0.8868	0.9280	0.9341
Using β'										
	0	0.8700	0.8994	0.8960	0.9123	0.9444	0.9488	0.9524	0.9777	0.9859
	1	0.8443	0.9004	0.8750	0.8841	0.9449	0.9247	0.9374	0.9803	0.9674
	4	0.8318	0.8334	0.8330	0.8836	0.8964	0.8925	0.9353	0.9519	0.9585
Using $\hat{k}_\beta$										
	0	0.8706	0.9030	0.9063	0.9133	0.9475	0.9536	0.9551	0.9795	0.9887
	1	0.8624	0.9184	0.8922	0.9064	0.9585	0.9384	0.9554	0.9865	0.9759
	4	0.8931	0.8932	0.8845	0.9281	0.9402	0.9322	0.9643	0.9760	0.9811

Table D.12: Attained service level for Figure 4.14.

L	ρ	$\alpha = 0.90$ $t=4$	$t=8$	$t=12$	$\alpha = 0.95$ $t=4$	$t=8$	$t=12$	$\alpha = 0.99$ $t=4$	$t=8$	$t=12$
0	$\frac{1}{2}$	1.4562	1.1883	1.1250	1.7770	1.3169	1.2169	2.7866	1.6461	1.3977
	1	1.2208	1.0926	1.0644	1.3544	1.1480	1.1013	1.7590	1.3144	1.2274
	2	1.1110	1.0470	1.0300	1.2008	1.0755	1.0488	1.4242	1.1779	1.1091
	4	1.0651	1.0220	1.0186	1.1144	1.0433	1.0264	1.2806	1.0930	1.0603
	6	1.0487	1.0174	1.0112	1.0871	1.0337	1.0187	1.2152	1.0658	1.0470
	8	1.0382	1.0114	1.0117	1.0765	1.0259	1.0144	1.1793	1.0710	1.0375
	10	1.0356	1.0106	1.0088	1.0676	1.0236	1.0143	1.1708	1.0686	1.0317
$\frac{1}{2}$	$\frac{1}{2}$	1.6414	1.2726	1.1674	1.9821	1.3997	1.2599	3.2745	1.7831	1.4958
	1	1.3045	1.1433	1.0901	1.4575	1.2134	1.1444	1.8399	1.3682	1.2430
	2	1.1577	1.0690	1.0450	1.2440	1.1072	1.0673	1.4943	1.2014	1.1278
	4	1.0932	1.0397	1.0278	1.1484	1.0664	1.0390	1.2992	1.1172	1.0779
	6	1.0662	1.0296	1.0231	1.1122	1.0463	1.0308	1.2402	1.0951	1.0544
	8	1.0530	1.0227	1.0151	1.0960	1.0410	1.0266	1.1956	1.0803	1.0528
	10	1.0464	1.0217	1.0146	1.0792	1.0296	1.0222	1.1684	1.0664	1.0475
1	$\frac{1}{2}$	1.7432	1.3281	1.2135	2.1574	1.4774	1.3054	3.3868	1.8762	1.5254
	1	1.3485	1.1665	1.1089	1.5134	1.2330	1.1531	1.9991	1.3972	1.2734
	2	1.1928	1.0857	1.0616	1.2894	1.1292	1.0917	1.5150	1.2187	1.1496
	4	1.1095	1.0521	1.0357	1.1684	1.0822	1.0496	1.3194	1.1387	1.0918
	6	1.0844	1.0408	1.0253	1.1284	1.0523	1.0382	1.2436	1.1013	1.0708
	8	1.0697	1.0319	1.0220	1.1083	1.0434	1.0332	1.2079	1.0867	1.0553
	10	1.0584	1.0275	1.0187	1.0887	1.0376	1.0263	1.1752	1.0764	1.0485
3	$\frac{1}{2}$	2.0505	1.4391	1.2824	2.5880	1.6369	1.3853	4.4704	2.0870	1.6704
	1	1.4893	1.2268	1.1562	1.6934	1.3166	1.2177	2.2799	1.5306	1.3502
	2	1.2659	1.1305	1.0838	1.3567	1.1788	1.1164	1.6381	1.2960	1.1933
	4	1.1553	1.0766	1.0569	1.2171	1.1099	1.0721	1.3659	1.1705	1.1179
	6	1.1204	1.0585	1.0398	1.1674	1.0784	1.0525	1.2878	1.1287	1.0867
	8	1.1000	1.0496	1.0357	1.1352	1.0645	1.0465	1.2332	1.1134	1.0771
	10	1.0869	1.0409	1.0301	1.1187	1.0588	1.0384	1.2040	1.0996	1.0683
6	$\frac{1}{2}$	2.2445	1.5189	1.3463	2.9569	1.7501	1.4661	5.2085	2.3673	1.8067
	1	1.5962	1.2822	1.1893	1.8406	1.3846	1.2594	2.4657	1.6115	1.4102
	2	1.3238	1.1657	1.1102	1.4473	1.2242	1.1500	1.7320	1.3566	1.2400
	4	1.1923	1.1019	1.0690	1.2687	1.1368	1.0957	1.4206	1.2118	1.1485
	6	1.1495	1.0786	1.0541	1.1958	1.1051	1.0767	1.3255	1.1593	1.1110
	8	1.1197	1.0640	1.0463	1.1650	1.0877	1.0619	1.2744	1.1373	1.0930
	10	1.1057	1.0576	1.0388	1.1445	1.0768	1.0519	1.2248	1.1184	1.0820

Table D.13: Values corrections for Figure 4.5.

L	ρ	$\beta = 0.90$ $t=4$	$t=8$	$t=12$	$\beta = 0.95$ $t=4$	$t=8$	$t=12$	$\beta = 0.99$ $t=4$	$t=8$	$t=12$
0	$\frac{1}{2}$	2.1065	1.5378	1.3422	2.5853	1.6504	1.4471	3.7405	2.1279	1.6209
	1	1.3631	1.1907	1.1380	1.4859	1.2372	1.1805	1.8366	1.3866	1.2855
	2	1.0977	1.0466	1.0324	1.1594	1.0727	1.0594	1.3352	1.1692	1.1028
	4	0.9809	0.9834	0.9928	1.0027	0.9979	0.9971	1.1269	1.0437	1.0365
	6	0.9439	0.9683	0.9810	0.9711	0.9768	0.9833	1.0631	1.0119	1.0031
	8	0.9323	0.9605	0.9718	0.9514	0.9676	0.9805	1.0182	0.9925	0.9875
	10	0.9206	0.9560	0.9703	0.9393	0.9640	0.9711	1.0010	0.9857	0.9814
$\frac{1}{2}$	$\frac{1}{2}$	2.1229	1.5078	1.3441	2.6210	1.6866	1.4471	4.0146	2.0935	1.6909
	1	1.4043	1.2038	1.1420	1.5721	1.2721	1.1925	1.9905	1.4798	1.3272
	2	1.1579	1.0860	1.0544	1.2242	1.1058	1.0766	1.4445	1.2073	1.1355
	4	1.0428	1.0238	1.0158	1.0776	1.0322	1.0241	1.2104	1.0867	1.0657
	6	1.0043	0.9970	0.9977	1.0250	1.0113	1.0041	1.1154	1.0451	1.0247
	8	0.9869	0.9854	0.9932	1.0003	0.9947	0.9950	1.0781	1.0295	1.0119
	10	0.9737	0.9841	0.9882	0.9891	0.9871	0.9949	1.0428	1.0074	1.0076
1	$\frac{1}{2}$	2.1626	1.5107	1.3296	2.5951	1.6396	1.4611	4.3824	2.1214	1.6948
	1	1.4522	1.2251	1.1523	1.6037	1.3012	1.2000	2.1387	1.4912	1.3267
	2	1.1861	1.0956	1.0668	1.2774	1.1377	1.0911	1.4948	1.2398	1.1518
	4	1.0730	1.0351	1.0240	1.1237	1.0593	1.0396	1.2447	1.1086	1.0622
	6	1.0391	1.0145	1.0112	1.0694	1.0288	1.0184	1.1665	1.0629	1.0388
	8	1.0141	1.0050	1.0037	1.0350	1.0133	1.0112	1.1065	1.0444	1.0268
	10	1.0014	1.0005	0.9983	1.0218	1.0072	1.0038	1.0826	1.0345	1.0150
3	$\frac{1}{2}$	2.2784	1.5300	1.3515	2.8426	1.7342	1.4719	4.8252	2.2884	1.7663
	1	1.5369	1.2676	1.1638	1.7518	1.3560	1.2290	2.3281	1.5601	1.3678
	2	1.2678	1.1363	1.0889	1.3702	1.1864	1.1285	1.6694	1.2871	1.2013
	4	1.1431	1.0700	1.0481	1.1978	1.0998	1.0673	1.3279	1.1677	1.1151
	6	1.0971	1.0484	1.0331	1.1343	1.0695	1.0484	1.2339	1.1160	1.0764
	8	1.0741	1.0388	1.0272	1.1034	1.0528	1.0362	1.1853	1.0921	1.0590
	10	1.0614	1.0306	1.0214	1.0826	1.0445	1.0280	1.1531	1.0758	1.0520
6	$\frac{1}{2}$	2.4791	1.6024	1.3836	3.0792	1.8474	1.5171	5.5196	2.4841	1.8589
	1	1.6163	1.2960	1.1982	1.8605	1.4093	1.2821	2.5789	1.6308	1.4432
	2	1.3249	1.1697	1.1127	1.4489	1.2292	1.1537	1.7359	1.3598	1.2367
	4	1.1924	1.1000	1.0687	1.2495	1.1330	1.0920	1.4007	1.1969	1.1368
	6	1.1359	1.0727	1.0502	1.1815	1.0969	1.0697	1.2902	1.1488	1.1052
	8	1.1102	1.0591	1.0411	1.1468	1.0807	1.0556	1.2353	1.1184	1.0813
	10	1.0919	1.0494	1.0341	1.1231	1.0671	1.0462	1.2000	1.1031	1.0689

Table D.14: Values corrections for Figure 4.10.

Appendix E

Derivations Chapter 5

E.1 Fitting a mixed-Erlang distribution

If X is a mixed-Erlang distributed variable, its pdf ($f(x)$) is the mixture of the pdfs of two Erlang distributed variables:

$$f(x) = p_1 \mu_1^{k_1} \frac{x^{k_1-1}}{(k_1-1)!} e^{-\mu_1 x} + p_2 \mu_2^{k_2} \frac{x^{k_2-1}}{(k_2-1)!} e^{-\mu_2 x}.$$

Hence, such a distribution has six parameters: p_1, p_2, k_1, k_2, μ_1, and μ_2. These parameters are nonnegative, $p_1 + p_2 = 1$, and $k_1, k_2 \in \mathbb{N}$.

Note that the expected value and variance of X are

$$\mathbb{E}[X] = p_1 \tfrac{k_1}{\mu_1} + p_2 \tfrac{k_2}{\mu_2},$$

and

$$\mathbb{V}[X] = p_1 \tfrac{k_1}{\mu_1^2}(1+k_1) + p_2 \tfrac{k_2}{\mu_2^2}(1+k_2) - p_1^2 \tfrac{k_1^2}{\mu_1^2} - p_2^2 \tfrac{k_2^2}{\mu_2^2} - 2p_1 p_2 \tfrac{k_1 k_2}{\mu_1 \mu_2}.$$

The coefficient of variation of X is determined using $\mathbb{C}_X = \sqrt{\mathbb{V}[X]}/\mathbb{E}[X]$. Since we have five unknown demand parameters ($p_2 = 1 - p_1$), we would need the first five moments to fit a mixed-Erlang distribution. Methods exist that use only the first two moments, see Tijms (1994), Van der Heijden (1993), and Janssen (1998). In Chapter 5 the method described by Janssen (1998) is used to fit a mixed-Erlang distribution to the first two moments:

$$
\begin{array}{ll|ll}
\multicolumn{2}{c|}{\text{if } \mathbb{C}_X < 1} & \multicolumn{2}{c}{\text{if } \mathbb{C}_X \geq 1} \\
\hline
k_1 = & \left\lfloor \frac{1}{\mathbb{C}_X^2} \right\rfloor & k_1 = & 1 \\
k_2 = & k_1 + 1 & k_2 = & 1 \\
p_1 = & \frac{k_2 \mathbb{C}_X^2 - \sqrt{k_2(1+\mathbb{C}_X^2) - k_2^2 \mathbb{C}_X^2}}{1+\mathbb{C}_X^2} & \mu_1 = & \frac{2}{\mathbb{E}[X]} \left(1 + \sqrt{\frac{\mathbb{C}_X^2 - \frac{1}{2}}{\mathbb{C}_X^2 + 1}}\right) \\
p_2 = & 1 - p_1 & \mu_2 = & \frac{4}{\mathbb{E}[X]} - \mu_1 \\
\mu_1 = & \frac{k_2 - p_1}{\mathbb{E}[X]} & p_1 = & \frac{\mu_1 (\mu_2 \mathbb{E}[X] - 1)}{\mu_2 - \mu_1} \\
\mu_2 = & \mu_1 & p_2 = & 1 - p_1.
\end{array}
\tag{E.1}
$$

We use $\boldsymbol{\zeta}$ to denote all the parameters of the mixed-Erlang distribution, so

$$\boldsymbol{\zeta} = [p_1, p_2, k_1, k_2, \mu_1, \mu_2].$$

Let D_ℓ denote the demand during ℓ periods. We assume that D_R can be accurately fitted to a mixed Erlang distribution with parameters $\boldsymbol{\zeta} = [p_1, p_2, k_1, k_2, \mu_1, \mu_2]$ and D_L to a mixed Erlang distribution with parameters $\boldsymbol{\eta} = [q_1, q_2, l_1, l_2, \lambda_1, \lambda_2]$.

E.2 Expected backlog at start replenishment cycle

This appendix provides the derivation of the expected backlog at the start of the replenishment cycle ($\mathbb{E}[(D_L - S)^+]$); it is based on Janssen (1998). We assume that the demand during lead time is mixed Erlang distributed with parameters q_1, q_2, l_1, l_2, λ_1, and λ_2. Furthermore, let $h_{l,\lambda}(x)$ denote the pdf of an Erlang distributed variable and $H_{l,\lambda}(x)$ its cdf. We have closed form expressions for these, namely:

$$
\begin{aligned}
h_{l,\lambda}(x) &= \lambda^l \frac{x^{l-1}}{(l-1)!} e^{-\lambda x}, \\
H_{l,\lambda}(x) &= 1 - e^{-\lambda x} \sum_{t=0}^{l-1} \frac{(\lambda x)^t}{t!}.
\end{aligned}
$$

If S is assumed to be fixed and $f_\eta(x)$ denotes the pdf of D_L, we can determine the expected backlog as follows:

$$\begin{aligned}
&\mathbb{E}\left[(D_L - S)^+\right] \\
&\quad = \int_S^\infty (x - S) f_\eta(x) dx \\
&\quad = \int_S^\infty (x - S) \sum_{j=1}^{2} q_j \lambda_j^{l_j} \frac{x^{l_j - 1}}{(l_j - 1)!} e^{-\lambda_j x} dx \\
&\quad = \sum_{j=1}^{2} q_j \int_S^\infty (x - S) \lambda_j^{l_j} \frac{x^{l_j - 1}}{(l_j - 1)!} e^{-\lambda_j x} dx \\
&\quad = \sum_{j=1}^{2} q_j \left(\int_S^\infty x \lambda_j^{l_j} \frac{x^{l_j - 1}}{(l_j - 1)!} e^{-\lambda_j x} dx - S \int_S^\infty \lambda_j^{l_j} \frac{x^{l_j - 1}}{(l_j - 1)!} e^{-\lambda_j x} dx \right) \\
&\quad = \sum_{j=1}^{2} q_j \left(\int_S^\infty \lambda_j^{l_j} \frac{x^{l_j}}{(l_j - 1)!} e^{-\lambda_j x} dx - S(1 - H_{l_j,\lambda_j}(S)) \right) \\
&\quad = \sum_{j=1}^{2} q_j \left(\int_S^\infty \lambda_j^{l_j+1} \frac{x^{l_j}}{l_j!} e^{-\lambda_j x} \frac{l_j!}{\lambda_j (l_j - 1)!} dx - S(1 - H_{l_j,\lambda_j}(S)) \right) \\
&\quad = \sum_{j=1}^{2} q_j \left(\frac{l_j}{\lambda_j} \int_S^\infty \lambda_j^{l_j+1} \frac{x^{l_j}}{l_j!} e^{-\lambda_j x} dx - S(1 - H_{l_j,\lambda_j}(S)) \right) \\
&\quad = \sum_{j=1}^{2} q_j \left(\frac{l_j}{\lambda_j} (1 - H_{l_j+1,\lambda_j}(S)) - S(1 - H_{l_j,\lambda_j}(S)) \right) \\
&\quad = \sum_{j=1}^{2} q_j \left(\frac{l_j}{\lambda_j} \sum_{t=0}^{l_j} \frac{(\lambda_j S)^t}{t!} e^{-\lambda_j S} - S \sum_{t=0}^{l_j - 1} \frac{(\lambda_j S)^t}{t!} e^{-\lambda_j S} \right).
\end{aligned}$$

Note that it is a closed form expression containing sums with a finite number of terms.

If S is assumed to be random, it can be fitted to a mixed Erlang distribution with parameters w_1, w_2, m_1, m_2, ρ_1, and ρ_2; the pdf that belongs to this mixed Erlang distribution is denoted by $f_\xi(x)$. The derivation is as follows:

$$\begin{aligned}
&\mathbb{E}\left[(D_L - S)^+\right] \\
&\quad = \int_0^\infty \int_z^\infty (x - z) f_\eta(x) dx f_\xi(z) dz \\
&\quad = \int_0^\infty \int_z^\infty (x - z) f_\eta(x) f_\xi(z) dx dz
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \int_z^\infty (x-z) \sum_{j=1}^{2} q_j \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} \sum_{i=1}^{2} w_i \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} dx dz \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \int_0^\infty \int_z^\infty (x-z) \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} dx dz \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \int_z^\infty (x-z) \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} dx dz \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \int_z^\infty x \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} dx dz \right. \\
&\quad \left. - \int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \int_z^\infty z \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} dx dz \right) \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \frac{l_j!}{\lambda_j (l_j-1)!} \int_z^\infty \lambda_j^{l_j+1} \frac{x^{l_j}}{l_j!} e^{-\lambda_j x} dx dz \right. \\
&\quad \left. - \int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} z \int_z^\infty \lambda_j^{l_j} \frac{x^{l_j-1}}{(l_j-1)!} e^{-\lambda_j x} dx dz \right) \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \frac{l_j}{\lambda_j} (1 - H_{l_j+1,\lambda_j}(z)) dz \right. \\
&\quad \left. - \int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} z (1 - H_{l_j,\lambda_j}(z)) dz \right) \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} \frac{l_j}{\lambda_j} \sum_{t=0}^{l_j} \frac{(\lambda_j z)^t}{t!} e^{-\lambda_j z} dz \right. \\
&\quad \left. - \int_0^\infty \rho_i^{m_i} \frac{z^{m_i-1}}{(m_i-1)!} e^{-\rho_i z} z \sum_{t=0}^{l_j-1} \frac{(\lambda_j z)^t}{t!} e^{-\lambda_j z} dz \right) \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\sum_{t=0}^{l_j} \frac{l_j \rho_i^{m_i} \lambda_j^t}{\lambda_j (m_i-1)! t!} \int_0^\infty z^{m_i-1} e^{-\rho_i z} z^t e^{-\lambda_j z} dz \right. \\
&\quad \left. - \sum_{t=0}^{l_j-1} \frac{\rho_i^{m_i} \lambda_j^t}{(m_i-1)! t!} \int_0^\infty z^{m_i-1} e^{-\rho_i z} z z^t e^{-\lambda_j z} dz \right) \\
&= \sum_{j=1}^{2} \sum_{i=1}^{2} q_j w_i \left(\sum_{t=0}^{l_j} \frac{l_j \rho_i^{m_i} \lambda_j^{t-1}}{(m_i-1)! t!} \int_0^\infty z^{m_i+t-1} e^{-(\lambda_j+\rho_i) z} dz \right. \\
&\quad \left. - \sum_{t=0}^{l_j-1} \frac{\rho_i^{m_i} \lambda_j^t}{(m_i-1)! t!} \int_0^\infty z^{m_i+t} e^{-(\lambda_j+\rho_i) z} dz \right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{j=1}^{2}\sum_{i=1}^{2} q_j w_i \\
&\left(\sum_{t=0}^{l_j} \frac{l_j \rho_i^{m_i} \lambda_j^{t-1}}{(m_i-1)!t!} \frac{(m_i+t-1)!}{(\lambda_j+\rho_i)^{m_i+t}} \int_0^\infty (\lambda_j+\rho_i)^{m_i+t} \frac{z^{m_i+t-1}}{(m_i+t-1)!} e^{-(\lambda_j+\rho_i)z} dz \right. \\
&\left. - \sum_{t=0}^{l_j-1} \frac{\rho_i^{m_i} \lambda_j^{t}}{(m_i-1)!t!} \frac{(m_i+t)!}{(\lambda_j+\rho_i)^{m_i+t+1}} \int_0^\infty (\lambda_j+\rho_i)^{m_i+t+1} \frac{z^{m_i+t}}{(m_i+t)!} e^{-(\lambda_j+\rho_i)z} dz \right) \\
&= \sum_{j=1}^{2}\sum_{i=1}^{2} q_j w_i \left(\sum_{t=0}^{l_j} \frac{l_j \rho_i^{m_i} \lambda_j^{t-1}}{(m_i-1)!t!} \frac{(m_i+t-1)!}{(\lambda_j+\rho_i)^{m_i+t}} \int_0^\infty h_{m_i+t,\lambda_j+\rho_i}(z) dz \right. \\
&\left. - \sum_{t=0}^{l_j-1} \frac{\rho_i^{m_i} \lambda_j^{t}}{(m_i-1)!t!} \frac{(m_i+t)!}{(\lambda_j+\rho_i)^{m_i+t+1}} \int_0^\infty h_{m_i+t+1,\lambda_j+\rho_i}(z) dz \right) \\
&= \sum_{j=1}^{2}\sum_{i=1}^{2} q_j w_i \left(\sum_{t=0}^{l_j} \frac{l_j \rho_i^{m_i} \lambda_j^{t-1}(m_i+t-1)!}{(m_i-1)!t!(\lambda_j+\rho_i)^{m_i+t}} \right. \\
&\left. - \sum_{t=0}^{l_j-1} \frac{\rho_i^{m_i} \lambda_j^{t}(m_i+t)!}{(m_i-1)!t!(\lambda_j+\rho_i)^{m_i+t+1}} \right).
\end{aligned}
$$

Note that again we have a closed form expression containing sums with a finite number of terms.

E.3 Distribution of demand during lead time and review period

As stated in Section 5.2 the mixed Erlang distribution is a special case of the phase type distribution. Since the phase type distribution is not a well-known distribution, this distribution is introduced in short.

A phase type distributed variable X consist of the sum of a (possibly random) number of exponentially distributed variables, possibly with different scale parameters. The phase type distribution can easily be linked to a Markov process; X measures the time until absorption in the corresponding Markov process. For a thorough introduction in phase type distributions see Neuts (1981) and Latouche and Ramaswami (1999).

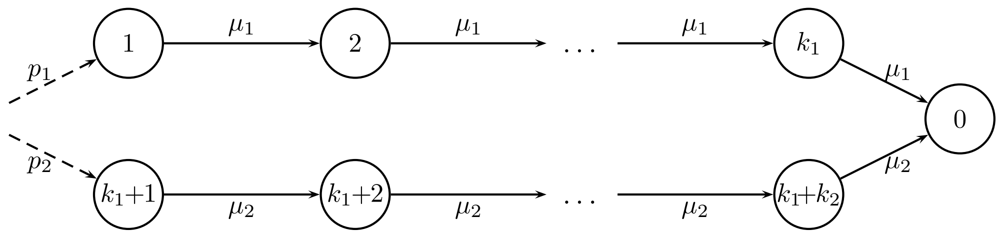


Figure E.1: The Markov process belonging to a mixed Erlang distribution.

Let us consider the mixed Erlang distribution. The corresponding Markov process is depicted in Figure E.1. The parameters p_1 and p_2 are the probabilities of drawing from either the first or the second Erlang distribution. The states $1, \ldots, k_1$ belong to that first distribution and the time it takes to get from state 1 to state 0 is the sum of k_1 exponentially distributed transition times with rate μ_1, so it is indeed Erlang distributed with shape parameter k_1 and scale parameter $1/\mu_1$. The states $k_1 + 1, \ldots, k_1 + k_2$ belong to the second Erlang distribution of the mixture of two Erlangs; the time it takes to get from state $k_1 + 1$ to 0 is the sum of k_2 exponentially distributed transition times with rate μ_2, so it is Erlang distributed with shape parameter k_2 and scale parameter $1/\mu_2$.

This Markov process is characterized by a $(k_1 + k_2) \times (k_1 + k_2)$ matrix T and a $(k_1 + k_2) \times 1$ vector $\boldsymbol{\tau}$, where:

$$T = \begin{array}{c}\left[\begin{array}{cccccc|cccccc}
-\mu_1 & \mu_1 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & -\mu_1 & \mu_1 & \ldots & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & -\mu_1 & \mu_1 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & -\mu_1 & 0 & 0 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & -\mu_2 & \mu_2 & 0 & \ldots & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & -\mu_2 & \mu_2 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & \ldots & -\mu_2 & \mu_2 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & -\mu_2
\end{array}\right]
\begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\\vdots\\0\\0\end{array}}\right\} k_1 \\ \left.\vphantom{\begin{array}{c}0\\0\\\vdots\\0\\0\end{array}}\right\} k_2 \end{array} \\
\underbrace{\hphantom{-\mu_1 \quad \mu_1 \quad 0 \quad \ldots \quad 0 \quad 0}}_{k_1} \qquad \underbrace{\hphantom{-\mu_2 \quad \mu_2 \quad 0 \quad \ldots \quad 0 \quad 0}}_{k_2}
\end{array} \tag{E.2}$$

and

$$\boldsymbol{\tau} = [\, \underbrace{p_1 \quad 0 \quad \dots \quad 0}_{k_1} \quad \underbrace{p_2 \quad 0 \quad \dots \quad 0}_{k_2} \,]. \tag{E.3}$$

The element T_{ij} in (E.2) is the transition rate from state i to state j and the element $\boldsymbol{\tau}_j$ in (E.3) denotes the probability to start in state j. According to Theorem 2.4.1 in Latouche and Ramaswami (1999) the density and distribution functions of the phase type distribution, characterized by T and $\boldsymbol{\tau}$, are

$$\begin{aligned} f_{T,\boldsymbol{\tau}}(x) &= \boldsymbol{\tau} \exp(Tx)\boldsymbol{t}, \\ F_{T,\boldsymbol{\tau}}(x) &= 1 - \boldsymbol{\tau} \exp(Tx)\mathbf{1}, \end{aligned}$$

where

$$\begin{aligned} \mathbf{1} &= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \\ \boldsymbol{t} &= -T\mathbf{1}, \end{aligned}$$

and

$$\exp(A) = \sum_{i=0}^{\infty} \frac{1}{i!} A^i.$$

Since both D_R and D_L are mixed Erlang distributed, they are phase type distributed with parameters T_R and $\boldsymbol{\tau}_R$ and with T_L and $\boldsymbol{\tau}_L$. According to Theorem 2.6.1 in Latouche and Ramaswami (1999) the sum of D_R and D_L, D_{R+L}, is again phase type distributed, with parameters

$$\begin{aligned} T_{R+L} &= \begin{bmatrix} T_R & \boldsymbol{t}_R \cdot \boldsymbol{\tau}_L \\ 0 & T_L \end{bmatrix}, \\ \boldsymbol{\tau}_{R+L} &= [\, \boldsymbol{\tau}_R \quad \tau_R^0 \boldsymbol{\tau}_L \,], \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{t}_R &= -T_R\mathbf{1}, \\ \tau_R^0 &= 1 - \boldsymbol{\tau}_R\mathbf{1}. \end{aligned}$$

Note that $\boldsymbol{\tau}_R\mathbf{1} = 1$, since $p_1 + p_2 = 1$, hence $\tau_R^0 = 0$.

E.4 Expected backlog at end replenishment cycle

If S is assumed to be fixed, the expected backlog at the end of the replenishment cycle is

$$\mathbb{E}\left[(D_{R+L}-S)^+\right]=\int_S^\infty (x-S)f_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx,$$

where $f_{T_{R+L},\boldsymbol{\tau}_{R+L}}$ denotes the pdf of a phase type distribution. We can rewrite this a little, so that we obtain

$$\begin{aligned}\mathbb{E}\left[(D_{R+L}-S)^+\right]&=\int_S^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx-\int_S^\infty Sf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx\\&=\int_S^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx-S\int_S^\infty f_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx\\&=\int_S^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx-S\big(1-F_{T_{R+L},\boldsymbol{\tau}_{R+L}}(S)\big).\end{aligned}$$

If S is random, we need to condition on the value of S and we obtain

$$\mathbb{E}\left[(D_{R+L}-S)^+\right]=\int_0^\infty\int_s^\infty (x-s)f_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dxf_{\boldsymbol{\xi}}(s)ds,$$

with $f_{\boldsymbol{\xi}}$ the pdf of a mixed Erlang distribution with parameters $\boldsymbol{\xi}$. Rewriting this expression leads to

$$\begin{aligned}\mathbb{E}\left[(D_{R+L}-S)^+\right]&=\int_0^\infty\Big(\int_s^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx-s\big(1-F_{T_{R+L},\boldsymbol{\tau}_{R+L}}(s)\big)\Big)f_{\boldsymbol{\xi}}(s)ds\\&=\int_0^\infty\Big(\int_s^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx\Big)f_{\boldsymbol{\xi}}(s)ds-\int_0^\infty sf_{\boldsymbol{\xi}}(s)ds\\&\quad+\int_0^\infty sF_{T_{R+L},\boldsymbol{\tau}_{R+L}}(s)f_{\boldsymbol{\xi}}(s)ds\\&=\int_0^\infty\Big(\int_s^\infty xf_{T_{R+L},\boldsymbol{\tau}_{R+L}}(x)dx\Big)f_{\boldsymbol{\xi}}(s)ds-\left(w_1\frac{m_1}{\rho_1}+w_2\frac{m_2}{\rho_2}\right)\\&\quad+\int_0^\infty sF_{T_{R+L},\boldsymbol{\tau}_{R+L}}(s)f_{\boldsymbol{\xi}}(s)ds.\end{aligned}$$

Note that both the cdf and pdf of a phase type distributed variable depend on $\exp(Tx)$, which is an infinite sum. Latouche and Ramaswami (1999) provide an algorithm to determine the value of $F_{T,\boldsymbol{\tau}}(x)$. We first need to define two variables, depending on the $n\times n$-matrix T:

$$\begin{aligned}c&=\max\{-T_{ii}\},\quad i\in\{1,\dots,n\},\\P&=\tfrac{1}{c}T+I,\end{aligned}$$

where I is the identity matrix. Note that c is the maximum of the negative of the diagonal of T and thus the diagonal elements of P are all less than or equal to one. The value of $F_{T,\boldsymbol{\tau}}(x)$ can now be determined numerically by the following algorithm (ϵ provides the precision):

$M := \boldsymbol{\tau}(I-P)^{-1}$;
$a_0 := \boldsymbol{\tau}\mathbf{1}$;
$k := 0$;
$\boldsymbol{v} := P\mathbf{1}$;
repeat
  $k := k+1$;
  $a_k := \boldsymbol{\tau}\boldsymbol{v}$;
  $\boldsymbol{v} := P\boldsymbol{v}$;
until $|\sum_{i=0}^{k} a_i - M| < \epsilon$;
$K_1 := k$;
for any x of interest **do**
  $p := e^{-cx}$
  $Y := pa_0$;
  for $k := 1$ **to** K_1 **do**
    $p := p\frac{cx}{k}$;
    $Y := Y + pa_k$;
  od
od
$F_{T,\boldsymbol{\tau}}(x) = 1 - Y$.

A similar algorithm can be provided for $f_{T,\boldsymbol{\tau}}(x)$.

Bibliography

Artto, K. A. and E. Pylkkänen (1999). An effective procedure for the distribution of magazines. *International Transactions in Operational Research*, 289–310.

Ashayeri, J., R. J. M. Heuts, H. G. L. Lansdaal, and L. W. G. Strijbosch (2006). Cyclic production-inventory planning and control in the pre-Deco industry: a case study. *International Journal of Production Economics*, 715–725.

Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science*, 1150–1160.

Azoury, K. S. and B. L. Miller (1984). A comparison of the optimal ordering levels of Bayesian and non-Bayesian inventory models. *Management Science*, 993–1003.

Bertsimas, D. and A. Thiele (2006). A robust optimization approach to inventory theory. *Operations Research*, 150–168.

Bookbinder, J. H. and A. E. Lordahl (1989). Estimation of inventory reorder level using the bootstrap statistical procedure. *IIE Transactions*, 302–312.

Bulinskaya, E. V. (1990). Inventory control in case of unknown demand distribution. *Engineering Costs and Production Economics*, 301–306.

Burgin, T. A. (1975). The gamma distribution and inventory control. *Operational Research Quarterly*, 507–525.

Evans, M., N. Hastings, and B. Peacock (2000). *Statistical distributions* (3 ed.). Wiley series in probability and statistics. New York: John Wiley & Sons, Inc.

Fricker, R. D. and C. A. Goodhart (2000). Applying a bootstrap approach for setting reorder points in military supply systems. *Naval Research Logistics*, 459–478.

Hayes, R. H. (1969). Statistical estimation problems in inventory control. *Management Science*, 686–701.

Heijden, M. C. van der (1993). *Performance analysis for reliability and inventory models*. Ph. D. thesis, Free University of Amsterdam.

Janssen, E., L. W. G. Strijbosch, and R. C. M. Brekelmans (2007). How to determine the order-up-to level when demand is gamma distributed with unknown parameters. Center discussion paper, Tilburg University.

Janssen, E., L. W. G. Strijbosch, and R. C. M. Brekelmans (2009). Assessing the effects of using demand parameters estimates in inventory control and improving the performance using a correction function. *International Journal of Production Economics*, 34–42.

Janssen, F. B. S. L. P. (1998). *Inventory management systems control and information issues*. Ph. D. thesis, Tilburg University.

Janssen, F. B. S. L. P., R. J. M. Heuts, and A. G. de Kok (1998). On the (R, s, Q) inventory model when demand is modelled as a compound Bernoulli process. *European Journal of Operational Research*, 423–436.

Kaplan, R. S. (1970). A dynamic inventory model with stochastic lead times. *Management Science*, 491–507.

Karmarkar, U. S. (1994). A robust forecasting technique for inventory and leadtime management. *Journal of Operations Management*, 45–54.

Kottas, J. F. and H.-S. Lau (1980). The use of versatile distribution families in some stochastic inventory calculations. *The Journal of the Operational Research Society*, 393–403.

Larson, C. E., L. J. Olson, and S. Sharma (2001). Optimal inventory policies when the demand distribution is not known. *Journal of Economic Theory*, 281–300.

Latouche, G. and V. Ramaswami (1999). *Introduction to matrix analytic methods in stochastic modeling*. ASA-SIAM series on statistics and applied probability. Philadelphia, USA: ASA SIAM.

Lau, H.-S. and A. H.-L. Lau (2003). Nonrobustness of the normal approximation of lead-time demand in a (Q, R) system. *Naval Research Logistics*, 149–166.

Liyanage, L. H. and J. F. Shanthikumar (2005). A practical inventory policy using operational statistics. *Operations Research Letters*, 341–348.

Lordahl, A. E. and J. H. Bookbinder (1994). Order-statistic calculation, costs and service in an (s, Q) inventory system. *Naval Research Logistics*, 81–97.

Lu, X., J.-S. Song, and A. Regan (2006). Inventory planning with forecast updates: approximate solutions and cost error bounds. *Operations Research*, 1079–1097.

Mittelhammer, R. C. (1996). *Mathematical statistics for economics and business.* New York: Springer-Verlag.

Neuts, M. F. (1981). *Matrix-geometric solutions in stochastic models.* Johns Hopkins series in the mathematical sciences. Baltimore, USA: The Johns Hopkins University Press.

Sani, B. and B. G. Kingsman (1997). Selecting the best periodic inventory control and demand forecasting methods for low demand items. *Journal of the Operational Research Society*, 700–713.

Scarf, H. (1958). A min-max solution of an inventory problem. In K. J. Arrow, S. Karlin, and H. E. Scarf (Eds.), *Studies in the mathematical theory of inventory and production*, pp. 201–209. Stanford, CA: Stanford University Press.

Segerstedt, A. (1994). Inventory control with variation in lead times, especially when demand is intermittent. *International Journal of Production Economics*, 365–372.

Silver, E. A., D. F. Pyke, and R. Peterson (1998). *Inventory management and production planning and scheduling* (Third ed.). New York: John Wiley & Sons.

Silver, E. A. and M. R. Rahnama (1986). The cost effects of statistical sampling in selecting the reorder point in a common inventory model. *Journal of the Operational Research Society*, 705–713.

Silver, E. A. and M. R. Rahnama (1987). Biased selection of the inventory reorder point when demand parameters are statistically estimated. *Engineering Costs and Production Economics*, 283–292.

Silver, E. A. and D. J. Robb (2008). Some insights regarding the optimal reorder period in periodic review inventory systems. *International Journal of Production Economics*, 354–366.

Snyder, R. D., A. B. Koehler, and J. K. Ord (2002). Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting*, 5–18.

Strijbosch, L. W. G. and R. M. J. Heuts (1992). Modelling (s, Q) inventory systems: parametric versus non-parametric approximations for the lead time demand distribution. *European Journal of Operational Research*, 86–101.

Strijbosch, L. W. G., R. M. J. Heuts, and E. H. M. van der Schoot (2000). A combined forecast–inventory control procedure for spare parts. *Journal of the Operational Research Society*, 1184–1192.

Strijbosch, L. W. G. and J. J. A. Moors (1999). Simple expressions for safety factors in inventory control. Center discussion paper, Tilburg University.

Strijbosch, L. W. G. and J. J. A. Moors (2005). The impact of unknown demand parameters on (R, S)-inventory control performance. *European Journal of Operational Research*, 805–815.

Strijbosch, L. W. G. and J. J. A. Moors (2006). Modified normal demand distributions in (R, S)-inventory control. *European Journal of Operational Research*, 201–212.

Strijbosch, L. W. G., J. J. A. Moors, and A. G. de Kok (1997). On the interaction between forecasting and inventory control. Few research memorandum, Tilburg University.

Strijbosch, L. W. G., A. A. Syntetos, J. E. Boylan, and E. Janssen (2010). On the interaction between forecasting and stock control: the case of non-stationary demand. *Accepted for publication in: International Journal of Production Economics*.

Syntetos, A. A. and J. E. Boylan (2006). On the stock control performance of intermittent demand estimators. *International Journal of Production Economics*, 36–47.

Thomas, D. J. (2005). Measuring item fill-rate performance in a finite horizon. *Manufacturing & Service Operations Management*, 74–80.

Tijms, H. C. (1994). *Stochastic models: an algorithmic approach.* Chichester: John Wiley & Sons.

Tyworth, J. E. and L. O'Neill (1997). Robustness of the normal spproximation of lead-time demand in a distribution setting. *Naval Research Logistics*, 165–186.

Watson, R. B. (1987). The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *Journal of Operational Research Society*, 75–82.

Yano, C. A. and H. L. Lee (1995). Lot sizing with random yields: a review. *Operations Research*, 311–334.

Yeh, Q.-J. (1997). A practical implementation of gamma distribution to the reordering decision of an inventory control problem. *Production and Inventory Management Journal*, 51–57.

Zeng, A. Z. and J. C. Hayya (1999). The performance of two popular service measures on management effectiveness in inventory control. *International Journal of Production Economics*, 147–158.

Zipkin, P. H. (2000). *Foundations of inventory management.* Boston: McGraw-Hill Higher Education.

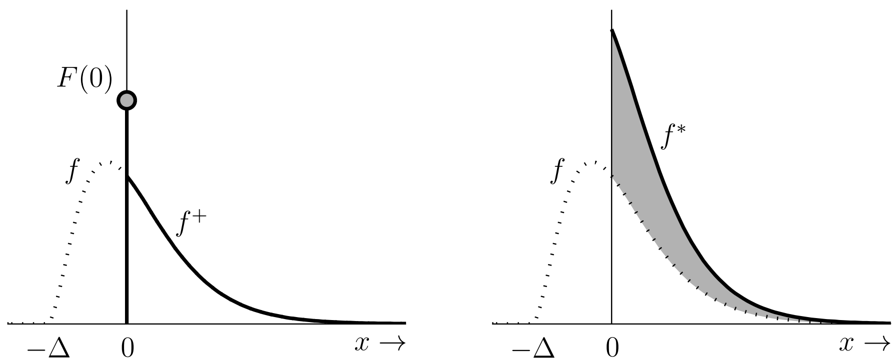
Nederlandse samenvatting

In het vakgebied voorraadbeheer gaat het erom om op een verstandige manier te kiezen wat we wanneer bestellen en hoeveel we dan bestellen. Om deze keuze te maken, wordt een aantal aannames gedaan omtrent het vraagproces, bijvoorbeeld over de kansverdeling van de vraag, de verwachte vraag en de spreiding van de vraag. Daarnaast wordt ook vaak a priori een voorraadbeleid vastgesteld. In dit proefschrift wordt uitgegaan van een periodieke controle van het voorraadniveau, waarna de voorraad wordt aangevuld tot een bepaald aanvulniveau. Er wordt dan bijvoorbeeld iedere dag, iedere week of iedere maand gecontroleerd hoe hoog de voorraad is en dan wordt een bestelling gedaan bij een toeleverancier die zodanig groot is dat het voorraadniveau weer op het aanvulniveau ligt. Deze bestelling wordt na een bepaalde levertijd geleverd. De grootte van de bestelling kan dus variëren, maar de tijd tussen bestellingen is altijd gelijk, namelijk de tijd tussen de periodieke controles (mits er vraag naar het product is geweest tussen twee periodieke controles). In het kort wordt dit beleid het (R, S) beleid genoemd, waarbij R de tijd tussen twee voorraadcontroles is en S het aanvulniveau; de levertijd wordt genoteerd door L. De waarde bij R wordt vaak van tevoren vastgesteld, bijvoorbeeld door afspraken met de toeleverancier; de waarde van S moet zodanig gekozen worden dat er genoeg voorraad is om aan de vraag van klanten te voldoen, maar aan de andere kant niet zo groot zijn dat er bergen voorraad in het magazijn liggen, want dat kost geld. Geld dat geïnvesteerd is in voorraden kan immers niet voor bijvoorbeeld productontwikkeling of een reclamecampagne gebruikt worden. De waarde van L hangt af van de toeleverancier en de afspraken die daarmee gemaakt zijn. Soms kan worden gesteld dat de levertijd gelijk is aan nul. Niet omdat de levering dan direct op de stoep staat, maar omdat een bestelling die na sluitingstijd gedaan wordt, de volgende dag voor opening aangeleverd zal zijn.

We willen het aanvulniveau dus zodanig kiezen dat de winst maximaal is (of dat de kosten minimaal zijn). Daarvoor moeten we weten wat het kost als een product niet in voorraad is wanneer een klant daarom vraagt. Maar wat doet een klant als het product waar hij om vraagt niet op voorraad is? Gaat hij dan het product ergens anders kopen en blijft hij dat dan ook doen? En de overige producten waarvoor de klant kwam, wil hij die wel kopen of doet hij dat dan ook bij een ander bedrijf? Is de klant bereid om te wachten totdat het product weer op voorraad is? Zo niet, kunnen we een spoedbestelling plaatsen waardoor we het product sneller kunnen leveren? En wat kost zo'n spoedbestelling dan? Dit soort overwegingen maakt het erg moeilijk om de precieze kosten van een nee-verkoop in te kunnen schatten.

Een andere manier om een goed voorraadniveau te kiezen is door de voorraad zo laag mogelijk te houden, maar wel hoog genoeg om aan een bepaald serviceniveau te voldoen. In dit proefschrift worden twee definities van service gebruikt. Ten eerste is dat de zogenoemde P_1-service: deze meet de fractie aanvulcycli waarin we aan alle vraag kunnen voldoen. Eén aanvulcyclus is de tijd tussen twee opeenvolgende leveringen. De tweede servicemaat is de zogenoemde P_2-service: deze meet de fractie van de vraag waaraan we direct kunnnen voldoen. Men kiest een bepaalde service, zoals "in 90% van de aanvulcycli moet aan alle vraag worden voldaan" ($\alpha = 0.90$) of "95% van de vraag moet direct geleverd kunnen worden" ($\beta = 0.95$), en vervolgens wordt S zodanig gekozen dat de gewenste service precies gehaald wordt.

In het eerste deel van dit proefschrift wordt gekeken naar nieuwe vraagverdelingen, namelijk twee modified verschoven gammaverdelingen. Deze verdelingen worden geconstrueerd door te beginnen met een verschoven gammaverdeling, waarbij negatieve realisaties mogelijk zijn; de cumulatieve verdelingsfunctie (cdf) van deze verdeling wordt genoteerd door F en de kansdichtheidsfunctie (pdf) door f. Een negatieve vraag is natuurlijk erg vreemd en we veranderen de verschoven gammaverdeling zodanig dat de negatieve realisaties verdwijnen. Je kunt dat doen door iedere negatieve realisatie de waarde nul te geven. Dat betekent dan dat er periodes zijn zonder vraag. De verdeling, genoteerd door F^+ (cdf) en f^+ (pdf), heeft een puntmassa op nul, waarvan de waarde dus gelijk is aan de kans op nul-vraag (genoteerd door $F(0)$), en is, voor $x > 0$ gelijk aan een verschoven gammaverdeling. Een tweede manier om de negatieve realisaties onmogelijk te maken is door ze simpelweg te negeren. In de verdeling die dan ontstaat, genoteerd door F^* (cdf) en f^* (pdf), wordt de kans op een negatieve realisatie in de verschoven gammaverdeling 'uitgesmeerd' over alle positieve realisaties. Zie ook de figuur bovenaan de volgende pagina voor een grafische weergave van beide kansdichtheidsfuncties. Deze nieuwe verdelingen zorgen ervoor dat vraagpatronen die min of meer verdeeld lijken te zijn volgens een gamma



(a): puntmassa op $x = 0$.

(b): afgekapt op $x = 0$.

Figuur: Kansdichtheidsfuncties van de modified verschoven gammaverdelingen.

verdeling op een meer flexibele manier geschat kunnen worden, aangezien deze nieuwe verdelingen een uitbreiding vormen op de gammaverdeling; de reguliere gammaverdeling is een speciaal geval van beide modified verschoven gammaverdelingen.

Als de vraag daadwerkelijk één van de twee modified verschoven gammaverdelingen volgt, is het mogelijk om het aanvulniveau te bepalen waarbij de P_1- of P_2-service gehaald wordt, als gebruik gemaakt wordt van een (R, S)-voorraadbeleid. Dit wordt uitgewerkt in paragraaf 2.4. Aangezien deze modified verschoven gammaverdelingen hier, voor zover bekend, voor het eerst afgeleid zijn, kan het goed zijn dat de verkeerde vraagverdeling gekozen wordt. Een reguliere of verschoven gammaverdeling ligt dan voor de hand, omdat de vraag toch min of meer gamma verdeeld lijkt. De verdelingsparameters van de reguliere gammaverdeling kunnen worden geschat met behulp van het eerste en tweede moment van de vraag; voor de verschoven gammaverdeling is ook het derde moment nodig. Als de reguliere of verschoven gammaverdeling ten onrechte wordt gebruikt om de vraag te modelleren, omdat de vraag verdeeld is volgens een modified verschoven gamma verdeling, zien we dat de behaalde service tussen ongeveer drie procentpunten onder de gewenste service tot drie procentpunten boven de gewenste service ligt. Dit verschil is niet heel groot, maar een behaalde service die drie procentpunten lager ligt dan de gewenste service kan toch een heleboel klanten, en dus omzet, kosten. Overigens is een hogere dan gewenste service ook niet zomaar goed; het impliceert immers dat er meer voorraad aangehouden wordt dan nodig is om de gewenste service te halen en dus dat er teveel geld geïnvesteerd wordt in voorraden.

Een belangrijk probleem binnen het vakgebied voorraadbeheer is dat de (verdeling van de) vraag onbekend is. Er zijn enkel wat historische observaties van de vraag en aan de hand daarvan zullen we belangrijke vraaggegevens moeten schatten. Een veelgebruikte methode is het aannemen van een bepaalde familie van verdelingen, bijvoorbeeld de normale of de gamma verdeling, en vervolgens de verdelingsparameters schatten aan de hand van de historische observaties. Merk op dat de schatters van de verdelingsparameters stochasten zijn en dat het aanvulniveau een functie is van de verdelingsparameters. Hierdoor is het aanvulniveau niet langer een deterministische parameter van het voorraadbeleid, maar een stochast en hierdoor wordt extra onzekerheid in het voorraadbeheer geïntroduceerd. De vraag is nu welke invloed deze extra onzekerheid heeft op de behaalde service. Intuïtief geldt dat meer onzekerheid aanleiding geeft tot het aanhouden van meer voorraad om deze onzekerheid op te kunnen vangen. Als er geen rekening wordt gehouden met de extra onzekerheid zal de behaalde service dus onder het gewenste niveau komen te liggen. In deel twee van dit proefschrift wordt analytisch (voor zover mogelijk) en met behulp van simulatie aangetoond dat dit inderdaad het geval is in het geval van normaal verdeelde, gamma verdeelde en mixed Erlang verdeelde vraag.

In hoofdstuk 3 wordt de normale verdeling besproken. Een belangrijke aanname is dat de vraag ook daadwerkelijk normaal verdeeld is, maar dat de verdelingsparameters onbekend zijn. Onder stricte aannames kunnen we analytisch aantonen dat de gewenste service niet gehaald wordt. In eerste instantie wordt aangenomen dat alleen de verwachte vraag (μ) onbekend is en geschat wordt met behulp van het steekproefgemiddelde. Zelfs in dit geval kunnen we aantonen dat zowel voor P_1 als P_2 de gewenste service niet gehaald wordt mits de gewenste service niet te laag is. Dit geldt echter voor alle gangbare gewenste serviceniveau's, dus dit levert geen beperkingen op in de praktijk. De behaalde service ligt lager als er minder historische waarnemingen zijn en als er meer onzekerheid in de vraag is; in beide gevallen zal de geschatte vraag immers meer onzekerheid bevatten. Ook het gewenste serviceniveau is van invloed, al is die invloed minder duidelijk. Men zal hiervoor naar het relatieve verschil tussen de gewenste en behaalde service moeten kijken; dan geldt dat hoe hoger de gewenste service, des te groter dit relatieve verschil is.

We vinden dankzij de theoretische afleiding ook een correctie: als we de standaarddeviatie in de functie van het aanvulniveau vervangen door de wortel van de voorspelfout, wordt de gewenste service weer gehaald.

Vervolgens wordt ook de spreiding van de vraag (σ^2) onbekend verondersteld;

deze wordt geschat met behulp van de steekproefvariantie. We kunnen aantonen dat het verwachte aanvulniveau lager is dan het aanvulniveau als alleen μ onbekend is. Hierbij gebruiken we de analytisch gevonden correctie en veronderstellen we nog wel dat de variatiecoëfficiënt ($\nu = \sigma/\mu$) bekend is. Een lager verwacht aanvulniveau betekent nog niet automatisch dat de gewenste service niet gehaald wordt, al is dat wel zeer waarschijnlijk, en als we simulaties uitvoeren om dit te controleren, blijkt de behaalde service behoorlijk onder het gewenste niveau te liggen.

Tenslotte nemen we ook aan dat de variatiecoëfficiënt onbekend is; op dit moment kunnen we geen analytische resultaten meer afleiden en zijn we aangewezen op simulatie. We gebruiken nog steeds de wortel van de geschatte voorspelfout in plaats van de (steekproek)standaarddeviatie. Na uitvoerige simulaties blijkt dat de behaalde service opnieuw lager uitvalt; de gemiddelde behaalde service ligt 2.03 procentpunt onder het gewenste niveau, maar er zijn uitschieters tot bijna 12 procentpunten.

Alleen het bepalen van de mate waarin de service niet gehaald is, is niet voldoende. Met behulp van simulatie en een nieuwe regressietechniek (geneste lineaire regressie genaamd) is een additieve correctiefunctie gevonden. Deze functie hangt af van het aantal historische waarnemingen, het gewenste serviceniveau en de onzekerheid in de vraag, gemeten aan de hand van de variatiecoëfficiënt van de vraag. Door deze correctiefunctie toe te passen met de werkelijke waarde van de variatiecoëfficiënt van de vraag, wordt het gewenste serviceniveau gehaald. Echter, in werkelijkheid kennen we deze waarde niet. Ook als we deze weer vervangen door de geschatte variatiecoëfficiënt ligt de behaalde service wel erg dicht bij de gewenste service, namelijk binnen één procentpunt (zie figuur 3.9).

Hoofdstuk 4 behandelt gamma verdeelde vraag, waarvan de verdelingsparameters onbekend zijn. Qua opzet lijkt dit hoofdstuk op hoofdstuk 3. Eerst wordt enkel de schaalparameter (θ) onbekend verondersteld. Deze wordt geschat met behulp van de bekende waarde van de vormparameter (ρ) en het steekproefgemiddelde. Het is aan te tonen dat deze schatter ook gamma verdeeld is en onder zeer stricte voorwaarden ($\rho = 1$ en $L = 0$) kunnen we dan laten zien dat de gewenste service niet gehaald wordt; zowel onder P_1-service als onder P_2-service. Als we deze voorwaarden wat verzachten, we staan een positieve levertijd toe en ρ is discreet, dan kunnen we, in geval van P_1-service, numeriek laten zien dat de behaalde service onder de gewenste service ligt voor gebruikelijke waarden van het gewenste serviceniveau. Indien we ook ρ volledig vrij laten, zijn er geen analytische resultaten meer mogelijk en in geval van de P_2-service zijn er überhaupt geen analytische resultaten meer mogelijk als we

$\rho = 1$ en $L = 0$ loslaten. In dit geval gebruiken we simulatie om te laten zien dat de behaalde service onder het gewenste niveau ligt. Verder zien we ook dat dit verschil groter is als ρ kleiner is, wat betekent dat de variatie in de vraag groter is. Het verschil is groter als het aantal historische waarnemingen kleiner is, wat betekent dat de variatie in de schatter van θ groter is. Het verschil is ook groter als de levertijd langer is, wat betekent dat we de vraag over een langere periode moeten voorspellen en dat zorgt weer voor meer onzekerheid. Tenslotte is het relatieve verschil groter als de gewenste service groter is en dat is ook weer niet geheel onlogisch, aangezien een hogere service moeilijker te behalen is.

Als ook de vormparameter onbekend wordt verondersteld, kunnen we enkel simulatie gebruiken om te laten zien dat de gewenste service opnieuw niet gehaald wordt als we schatters gebruiken in plaats van de werkelijke, maar onbekende, waarde van de verdelingparameters. Het verschil tussen de behaalde en gewenste service is bovendien groter dan wanneer enkel de schaalparameter onbekend is, dus we zien opnieuw dat meer onzekerheid een lagere behaalde service impliceert. Het gemiddelde verschil tussen de behaalde en gewenste service is 8.11 (P_1-service) en 8.20 (P_2-service) procentpunten, met uitschieters naar bijna 30 procentpunten.

Onder de zeer stricte voorwaarden dat $\rho = 1$ en $L = 0$ kunnen we een correctie vinden, waarmee de service onder die voorwaarden weer gehaald wordt. Uiteraard zal dit niet voldoende zijn als we meer onzekerheid toevoegen, maar het zou alvast een goede eerste stap kunnen zijn. En inderdaad, door deze eerste verbetering toe te passen ligt de behaalde service alweer dichterbij de gewenste service. Het gemiddelde verschil in geval van P_1-service is nu 5.13 procentpunten en in geval van P_2-service is dat 5.20 procentpunten; voor beide servicematen zijn er uitschieters tot ongeveer 25 procentpunten.

Aangezien het verschil tussen de behaalde en de gewenste service nog steeds aanzienlijk is, zoeken we ook hier een correctiefunctie met behulp van simulatie en geneste lineaire regressie. Deze functie is afhankelijk van de vormparameter van de verdeling (en daarmee ook van de variatiecoëfficiënt van de vraag), het aantal historische waarnemingen, de gewenste service en de levertijd. Als we de werkelijke, maar onbekende, waarde van de vormparameter gebruiken in de correctiefunctie, dan zal de gewenste service behaald worden. In de praktijk is dit echter niet mogelijk en zal deze weer vervangen moeten worden door een schatter van deze parameter. De gewenste service wordt nu niet meer in alle gevallen gehaald, maar de behaalde service ligt in ieder geval dicht bij de gewenste service. Het gemiddelde verschil is 0.91

procentpunten in geval van P_1-service en 1.24 procentpunten in geval van P_2-service, met uitschieters tot een kleine 8 procentpunten.

Tenslotte hebben we dit voorraadbeleid toegepast op werkelijke vraaggegevens. Uit deze case study blijkt dat als de gammaverdeling met geschatte verdelingsparameters zonder correcties wordt toegepast, de behaalde service ver onder het gewenste niveau ligt. De eerste verbetering helpt al enigszins en als daar bovenop de correctiefunctie gebruikt wordt, wordt de gewenste service bijna gehaald; in een enkel geval komt de behaalde service zelfs iets boven de gewenste service uit.

In hoofdstuk 5 onderzoeken we opnieuw het effect van het niet meenemen van onzekerheid op het behalen van de service. In dit geval kijken we niet naar wat nu die oorzaak is; we nemen wel meer bronnen van onzekerheid mee in de analyse. In hoofdstukken 3 en 4 zijn de aanvulniveaus onzeker, maar de tijd tussen twee voorraadcontroles en de lengte van de levertijd worden bekend verondersteld. Als ook deze onzeker zijn, zal dit weer extra onzekerheid toevoegen in het voorraadbeleid en, naar verwachting, zal de behaalde service weer lager zijn dan de gewenste service. In dit hoofdstuk zijn we uitgegaan van mixed-Erlang verdeelde vraag, vanwege de handige eigenschappen en de flexibiliteit van deze verdeling. Door die handige eigenschappen kunnen we analytisch laten zien dat de gewenste service inderdaad niet gehaald wordt als er geen rekening gehouden wordt met de extra onzekerheid in de tijd tussen twee voorraadcontroles, in de levertijd en in het aanvulniveau. Bovendien zien we ook dat hoe meer onzekerheid er is, hoe hoger de voorraad moet zijn om ervoor te zorgen dat de gewenste service gehaald wordt, precies zoals intuïtief ook wel te verwachten is. Het verschil tussen de gewenste en behaalde service kan oplopen tot meer dan 20 procentpunten.

Kortom, in deel II wordt aangetoond dat als (extra) onzekerheid in de beleidsparameters (R, S en L) genegeerd wordt, de gewenste service niet gehaald wordt en dat het verschil tussen de behaalde en gewenste service aanzienlijk kan zijn. Aan de andere kant zal de voorraad ook aanzienlijk hoger moeten zijn om al deze onzekerheid op te vangen; dit laatste wordt vooral geïllustreerd in de figuren 5.1–5.3. Het kan dus lonen om de oorzaak van de onzekerheid aan te pakken in plaats van de symptomen (te lage service) te bestrijden door een hogere voorraad aan te houden.

Author index

Subject index